DOCUMENT RESUME

ED 227 134 TM 830 126

AUTHOR Shepard, Lorrie A.
TITLE Assessment of Learning Disabilities.
INSTITUTION ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO ERIC-TM-84
PUB DATE Dec 82
CONTRACT 400-78-0003
NOTE 76p.
AVAILABLE FROM ERIC/TM, Educational Testing Service, Princeton, NJ 08541 ($6.00).
PUB TYPE Information Analyses - ERIC Information Analysis Products (071)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Educational Diagnosis; Elementary Secondary Education; *Learning Disabilities; *Learning Problems; *Mainstreaming; Political Issues; *Psychoeducational Methods; Social Problems; Student Evaluation; Student Needs; Student Placement; Test Use

ABSTRACT

The assessment and diagnosis of learning disabilities (LD) in the school is problematic. How do educators determine who is learning disabled? What practices are recommended? The main focus of the paper is on specific, relatively technical points that influence the validity of assessment. Since technical concerns are only one of the factors influencing the validity of placements, this paper is organized into two sections: the context of LD identification and technical issues in LD assessment. Specific propositions regarding the context of LD identification are advanced with supporting evidence: overidentification in the LD category; ambiguity in the definition of LD and local idiosyncratic criteria; students' needs for special help; parental demand and pressure from regular education; teaching and system failures; and the consequences of overidentification. Technical topics related to steps in the assessment process are referral bias; normal variability and clinicians' vertigo; technical adequacy of tests; specialists' knowledge of test adequacy and measurement concepts; significant ability-achievement discrepancy; interpreting subtest scatter; using age norms to evaluate processing deficits; behavorial indicators, informal assessments, and clinical hypothesis testing; and exclusion and bias. Recommendations made include contextual changes that are likely to help clinicians be willing to make more rigorous diagnoses and improved training and retraining of specialists. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227134

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ASSESSMENT OF LEARNING DISABILITIES

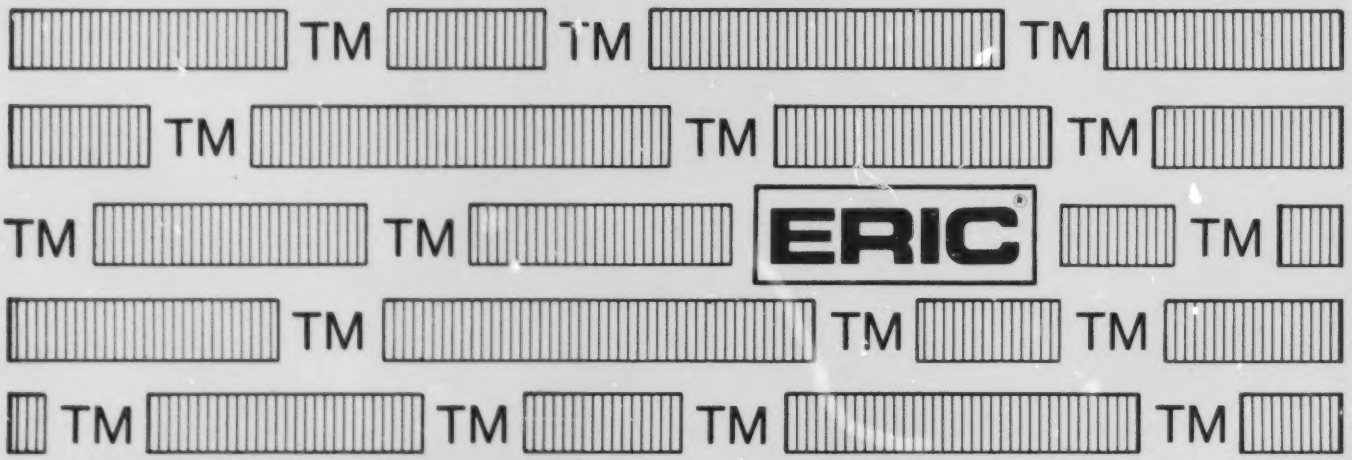
ERIC/TM REPORT 84

by
Lorrie A. Shepard

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B.M. Wilkerson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



TM 130 126

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08541

ERIC/TM Report 84

ASSESSMENT OF LEARNING DISABILITIES

by

Lorrie A. Shepard

University of Colorado

December 1982

ERIC Clearinghouse on Tests, Measurement, and Evaluation

Educational Testing Service, Princeton, NJ 08541

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U. S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the National Institute of Education.

ERIC Clearinghouse on Tests,
Measurement, and Evaluation
Educational Testing Service
Princeton, NJ 08541

TABLE OF CONTENTS

Introduction.....2

The Context of LD Identification.....6

Overidentification in the LD Category.....6

Ambiguity in the Definition of LD and Local Idiosyncratic Criteria.....8

Students' Needs for Special Help.....13

Parental Demand and Pressure from Regular Education.....15

Teaching and System Failures.....17

The Consequences of Overidentification.....18

Technical Issues in LD Assessment.....22

Referral Bias.....22

Normal Variability and Clinicians' Vertigo.....24

Technical Adequacy of Tests.....27

Specialists' Knowledge of Test Adequacy and Measurement Concepts.....31

Significant Ability-Achievement Discrepancy.....35

Interpreting Subtest Scatter.....39

Using Age Norms to Evaluate Processing Deficits.....42

Behavioral Indicators, Informal Assessment, and Clinical Hypothesis Testing.....44

Exclusion and Bias.....47

Summary and Recommendations.....52

References.....56

Appendix A.....62

Table 1.....64

Table 2.....68

Introduction

The assessment and diagnosis of learning disabilities (LD) is problematic. In educational and psychological measurement, difficulties are always encountered whenever observed signs and behaviors are used to infer a person's underlying characteristics. However, the problems in assessing learning disabilities are unusually serious because, in addition to ordinary technical problems, the construct is so poorly understood. Inadequate conceptualization leads to invalid measurement and misidentification, which creates its own vicious cycle. Researchers then study misidentified populations to try to deduce signs associated with the disorder and improve the conceptualization of the trait, but these efforts are doomed from the start because of the confounding of valid and invalid cases (see Harber, 1981; Kavale & Nye, 1981; or Olson & Mealor, 1981 for summaries of population definitions in LD research).

This report is not addressed to the researcher's difficulties in identifying LD, however, but rather to the practitioner's dilemma when trying to identify LD in the schools. Public Law 94-142 includes the learning disabled among the handicapped who are guaranteed the right to a free and appropriate education. Thus, educators are required to identify and serve a type of handicapped child that researchers have so far failed to define. Practitioners must proceed in making diagnoses despite the recognized difficulties. Moreover, the consequences of misidentification are much more serious for practitioners since their decisions are made about

individuals, whereas researchers make decisions about groups.[1]

The purpose of this report is to present a summary of the issues in LD assessment. How do educators determine who is learning disabled? What practices are recommended? The main focus of the paper is on specific, relatively technical points that influence the validity of assessment, such as the psychometric adequacy of tests, interpretation of test-score profiles, and the meaning of behavioral checklists. A basic premise is, however, that technical concerns are only one of the factors influencing the validity of placements. When a diagnosis is being made other forces, such as parental demand for special education services may become much more salient than the interpretation of a significant discrepancy score. Therefore, the paper is organized into two major sections: (1) the context of LD identification and (2) technical issues in LD assessment. In the first section, specific propositions regarding the context of LD identification are advanced with supporting evidence. For example, the first proposition is that children are being overidentified in the LD category. Some factors such as inadequate definition merely contribute to misidentification, i.e., either over or underidentification could result. Other factors, however, such as the lack of programmatic alteratives for children in need of remedial services, lead systematically to overidentification. The theoretical, social, and political problems discussed in the first section impinge occasionally on the technical points in the second section. Generally, whenever there is ambiguity in diagnostic evidence, other

---

[1] It is a guiding principle in educational and psychological measurement that technical requirements for reliability and validity depend upon test use. When tests or observational data are used to make individual decisions such as selection or classification, the technical standards are perforce much more stringent (Cronbach, 1970; Mehrens & Lehman, 1975) because it is recognized that errors in individual decisions have more serious consequences than when data are gathered only for research or institutional planning purposes.

pressures arguing for placement will have greater impact. It is also true that the technical problems themselves sometimes lead to misidentification (for example, random errors could be in either direction); but, in other instances it can be shown that the technical errors tend systematically to contribute to overidentification. Given the presenting problem of substantial overidentification in the LD category, it is helpful to distinguish between technical problems of the two types, those leading to random errors and those leading to systematic overidentification.

Finally, in the concluding section of the paper, recommendations are made for the improved training or retraining of specialists. What general perspectives and what specific technical competencies should professionals have to ensure the validity of LD placements? Once again, however, it is recognized that professional training alone is not likely to be effective in reducing overidentification in LD. After all, specialist surveys sometimes reveal that school psychologists or LD teachers knowingly misidentify slow learners as LD to obtain special servies. In such cases, the problem is not attributable to a faulty test or ignorance about test score statistics. Therefore, the recommendations include contextual changes that are likely to help clinicians be willing to make more rigorous diagnoses.

Terminology

It may be helpful in delimiting the purpose and scope of the paper to be clear on the meaning of key terms used in the foregoing paragraphs and in the remainder of the text. In special education, the process of identification includes the steps of referral, assessment, staffing, and placement. Thus, identification is more inclusive than assessment; it encompasses the entire process whereby children are determined to be LD and declared eligible for special education. Note that the first section of the

paper deals with the context of the entire identification process. The second section is addressed more specifically to technical issues in LD assessment. Although some other aspects of the process are discussed such as referral bias and team decision making, the focus is primarily on assessment.

Assessment is broadly defined to include both formal and informal data collection procedures for making educational decisions about individuals (Ysseldyke, 1977). Formal tests including standardized tests are merely one form of assessment. Although assessment is not confined to standardized testing, tests and test score interpretation receive the greatest attention here because thus far standardized tests have been the predominant source of evidence for LD diagnosis in the schools (Poland, et al., 1979; Thurlow, 1980; Thurlow & Ysseldyke, 1979). This reliance on test score information for LD identification is unlike assessment practices for some other categories of handicap, such as emotionally disturbed, where decisions are based primarily on non test evidence.

LD pupils may be assessed for many reasons including instructional planning and program evaluation (e.g., is special education effective?). The title of this report, however, "assessment of LD," refers only to assessment for the purpose of diagnosis, that is, for determining whether a child is or is not LD. Assessment practices to guide instruction are outside the scope of this report.

The Context of LD Identification

Overidentification in the LD Category

Learning disabilities is recognized to be a generic term that refers to a heterogeneous population (Hammill, et al., 1981; Weener & Senf, 1982). The assertion made here that children are being overidentified as LD does not presume that all cases must fit one, simplistic profile of LD or else be considered invalid. Rather, the claim is made that in addition to several legitimate subtypes of LD there are many children in the school LD population who should not be in this category by any definition or criteria. The "overidentified" cases include children with other handicaps and children who are behind in school but who are not handicapped. Poplin (1981) described the types of children likely to be mislabeled as LD, "In addition to the truly handicapped learning disabled person, we find the learning disability specialist serving students with behavior problems, students from different cultural backgrounds, slow learners, the poorly taught, and remedial education students" (p. 330).

Early research on the characteristics of LD populations contained only limited data such as IQ and achievement test scores, but was suggestive that (in the aggregate) the empirical results may not match theoretical definitions. For example, Kirk and Elkins (1975) surveyed Child Service Demonstration Centers and found that in half of the centers children had been classified as LD with IQs of 69 and below. Across all 21 states a disproportionate number of cases (35%) had IQs below 90. Nearly identical results were obtained by Norman and Zigmond (1980).

Recently, more complex investigation have been conducted to describe the characteristics of LD populations and to determine whether they can be differentiated from other low achievers. Ysseldyke, et al. (1982) compared LD and low achieving pupils from the same schools on 49 variables and found that the amount of overlap between the two groups was form 82% to 100%. Also, when the test scores of the two groups were examined in light of the federal definition of LD, 40% or more appeared to be misclassified. In a similar study, Warner, et al. (1980) found that although secondary LD pupils tended to be lower in both achievement and ability than non-special-education low achievers, they did not have greater IQ-achievement discrepancies; nor were the two groups different on any other variables tried in discriminant analyses.

Shepard and Smith (1981) used hierarchical analyses to assign 1,000 representative LD cases to identifiable subgroups. Eight different criteria were considered whereby children could be classified as legitimately LD, including significant ability-achievement discrepancy, combinations of weak signs, known brain injury, and clinical evidence of processing deficit. Cases were counted as validly LD if they satisfied any one of the criteria, but only 43% of the school LD population were accounted for by these subgroups. The remaining 57% of the cases included other handicaps (10%), non-English-dominant pupils (7%), slow learners (11%), and minor behavioral problems (4%), as well as normal children (disproportionately occuring in high socioeconomic status districts). The Shepard and Smith findings were corroborated by a similar large scale evaluation study conducted in California. Using a combination of characteristics to define LD pupils who were seriously disabled or at least "at risk" (normal IQ>85, achievement one-half year below grade level, and a significant verbal-performance IQ

discrepancy), the researchers concluded that only 4 out of 10 students currently in LD would qualify for special education (Craig, Myers, & Wujek, 1982).

In the following sections, the plausible causes of overidentification are summarized. Overidentification of LD can be attributed at least in part to ambiguity in the definition, psychometrically inadequate tests and lack of technical training of specialists, students' needs for special help, parental demand, pressure from regular education, and teaching failures.

Ambiguity in the Definition of LD and Local Idiosyncratic Criteria

The definition of learning disabilities has been controversial since the term was first popularized by Kirk (1963). LD was intended as a neutral, descriptive label for children who had previously been called brain-injured, neurologically impaired, perceptually handicapped or said to suffer minimal brain dysfunction. These prior constructs were themselves ambiguous since they inferred a neurological condition that cannot, by definition, be demonstrated. Cruickshank (1972) concluded that there "is no common denominator of understanding" (p. 382).

As was suggested in the introduction, there are many psychological constructs which are difficult to define precisely, e.g., self concept, intelligence. Usually, theory and concrete measurements are allowed to evolve together. However, when a construct is made a part of public policy the theory and conceptual development may be fixed at that point or at least seriously constrained. For a discussion of the governmental influence on the definition of LD see Weener and Senf (1982). When policy is being implemented, attention is focused on "operational" criteria rather than

conceptual understanding. Mercer, Forgnone and Wolking (1976) chronicled the proliferation of different definitions across the states. Sabatino and Miller (1980) concluded that divergence of definitions "has pushed local practitioners to develop disparate procedures in accordance with broadly specified state criteria" (p. 76).

The current legal definition was that adopted as part of PL 94-142:

> "Specific learning disability" means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage (U.S.O.E., 1977, p. 65083).

Although this definition, taken almost word for word from the National Advisory Committee on Handicapped Children (1968), represented the state of the art, it has many limitations and may even foster some misconceptions about the nature of LD. In an attempt to clarify the theoretical understanding of the LD construct, a new definition was proposed by a joint committee (NJCLD) of the American Speech-Language-Hearing Association, the Association for Children and Adults with Learning Disabilities (ACLD), the Council for Learning Disabilities, the Council for Exceptional Children's Division of Children with Communication Disorders, the International Reading Association, and the Orton Dyslexia Society. The following definition has now been adopted by all of the participating organizations except the ACLD.

> Learning disabilities is a generic term that refers to a heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning or mathematical abilities. These disorders are intrinsic to the individual and presumed to be due to central nervous system dysfunction. Even though a learning disability may occur concomitantly with other handicapping conditions (e.g., sensory impairment, mental retardation, social and emotional disturbance) or environmental influences (e.g., cultural differences, insufficient/inappropriate instruction, psychogenic factors), it is not the direct result of those conditions or influences (Hammill, Leigh, McNutt, & Larsen, 1981).

The rationale for each of the elements in the definition is explained in Hammill et al. (1981).

Although disagreement and ambiguity may be the features of the LD definition that have the most pervasive influence on assessment practices, there are positive conceptual threads in these theoretical statements which should guide the development of diagnostic techniques. That is, assessment techniques discussed in the second section of the paper should be evaluated in terms of their fidelity to the intended concepts in the definition. In this way, assessment practices can be guided more by theory rather than constrained by impoverished, oversimplified operational rules.

It is argued here that the key elements in defining LD continue to be difficulty in school learning, discrepancy (or anomaly) in cognitive functioning, the intrinsic nature of the disorder, and exclusion (or ruling out) of other primary causes. Learning problems are a necessary but not sufficient condition for the diagnosis of LD, since individuals could have trouble learning and not be considered LD if the failure was more accurately attributed to mental retardation, poor attendance of poor teaching.

Discrepancy. The concept of ability-achievement discrepancy was central to the definitions of LD posed by Bateman (1964) and Kirk and Bateman (1962). It is also the key operational criterion in the current federal guidelines. Because the concept of discrepancy is "the most widely accepted sign of learning disability" (Weener & Senf, 1982, p. 1060), some states and districts have tried to be strict about LD identification by imposing a formula for a statistically significant discrepancy. Unfortunately, such operational definitions seldom capture the richness of the conceptual definition (Senf, 1978). There are many reasons why such formulae will fail as the sole criterion for LD. For one, significant differences indicate underachievement of all types, not just LD. Conversely, a valid disability could depress the obtained IQ score and prevent a significant discrepancy score. Because discrepancy formulae have been applied simplistically there may be a tendency now to avoid them altogther. It should be remembered, however, that discrepant or anomalous cognitive functioning is essential to the concept of LD. Nelson Rockefeller is often cited as an example of someone with a learning disability and is an excellent example of the anomaly characteristic. He was a brilliant man who had trouble reading. He memorized his speeches or had them written in very large letters. This inability was "suprising" or discrepant given all the other evidence of his intellectual ability. Difficulty in reading would not be considered a sign of LD, however, if it were consistent with generally depressed intellectual functioning. The notion of anomalous intellectual performance was a part of the impetus originally for hypothesizing minimal brain damage. Other signs suggested that an individual had the necessary intellectual ability, but he or she failed on a particular type of task, leading to the inference of damage in a specific area just like the victims of stroke. From the beginning, the construct was intended to be distinct from low IQ.

Intrinsic disorder. The new definition includes a restatement of the basic understanding that learning disabilities are an internal characteristic of the individual. The attribution of cause to central nervous dysfunction does not mean that hard evidence of the dysfunction is required for diagnosis. Rather this is an amplification of the underlying construct. "the phrase is intended to spell out clearly the intent behind the statement that learning disabilities are instrinsic to the individual" (Hammill et al., 1981, p. 340). The definition of an intrinsic disorder is really equivalent to the exclusionary clause. That is, the disability is not imposed on the individual by external factors such as lack of opportunity to learn.

Exclusion. Although the exclusionary clause is sometimes mocked, e.g., Gallagher's (1976) reference to the "nonhorse" aspects of LD definitions, it is, in fact, customary to establish the discriminant validity of psychological constructs by saying how they are to be distinquished from similar and related traits. The exclusions in the definitions of LD are consistent with the concepts of discrepancy and the intrinsic nature of the disorder. If an individual is having serious learning problems because he or she is mentally retarded (and all evidence is consistent with the diagnosis of retardation), then there are no surprises or discrepancies, no reason to invoke the second construct of LD. Similarly if a first grader is from a very depressed socioeconomic background, has not had books at home, and has not been exposed to letter sounds in kindergarten, it is entirely within expectation that this child will be behind his or her classmates in learning to read. So long as progress is made on material presented at the child's own rate and so long as there is nothing anomalous in <u>how</u> the child learns, is no reason to posit a learning disability.

The exclusion clause has been attacked because, when overzealously interpreted, it appeared to deny the possibility that a poor child or a blind child could also be learning disabled. This is clearly not the case. In fact, there may be good reason to believe that conditions of extreme poverty leading to prenatal malnutrition or sensory deprivation could actually increase the incidence of the intrinsic disorders in the "excluded" populations. The important conceptual distinction hinges on the primary cause of the observed learning difficulty. In some ways it helps to think of LD as an add-on construct--is the child's problem greater than (more discrepant than) would be expected given other known sources of learning difficulty?

Students' Needs for Special Help

In the Shepard and Smith (1981) study, it was concluded that only about half of the children currently placed in LD programs in Colorado were legitimately identified as handicapped. Ten percent had other handicaps, e.g., were mentally retarded or emotionally disturbed, and 43% were legitimately labeled as LD. For the remaining half of the school-identified LD population who were not handicapped, Shepard and Smith developed two distinct hypotheses or explanations to account for the overidentification: (1) helping nonhandicapped children with special needs and (2) removing problem children from the regular classroom. The first explanation which we called a commendable motive, is the topic of this section. The second explanation, which is not so commendable, is addressed in a later section entitled teaching and system failures.

Based on in-depth readings of representative cases, Shepard and Smith concluded that over half of the misidentified pupils (approximately half of

the half or 30% of the total sample) did have serious problems in school and did need special help. These cases included children from Spanish language backgrounds and slow learners whose measured IQs were in the 75 to 90 range (but who also did not have any consistent clinical evidence of LD). Often the staffing minutes for these children included statements that the child had a significant discrepancy or had a processing deficit in an area such as auditory discrimination, even though the test results would not support these conclusions. We believe there were two quite different reasons for these misstatements. Sometimes the professionals were genuinely misinformed about how to compute discrepancies or interpret test scores. These problems are addressed in the technical portion of the paper. Just as often, however, professionals quoted the litany of "significant discrepancy" because it is in the letter of the LD guidelines, but they in fact believed that the child should be placed because he or she needed the benefit of one-to-one instruction. This practice is obvious when the staffing summaries for all the cases in one district read like carbon copies of each other. Obviously, errors of this type will not be corrected by increased sophistication regarding assessment.

The trend to place children who need help regardless of whether they are really LD appears to be greatest when alternative programs are not available and when there are no disincentives for increasing the total number of LD. Directors of special education in some rural districts testified, for example, that overidentification is likely to occur in LD because it is the only recourse for help and because the label is both vague and nonstigmatizing. The well intentioned desire to provide services to children in need corresponds to what Hewett and Forness (1974) called the service motivations of professionals as opposed to scientific motivations.

It also corresponds to what is termed in the field the Statue of Liberty effect, "Give me your tired, your poor, etc." Because of their earnest desire to help children (and also to appear helpful and omniscient to their regular education colleagues), special education professionals believe they should take all comers.

Parental Demand and Pressure from Regular Education

Added to uncertainty and ambiguity regarding the diagnosis of LD and the professional's own desire to help children are the pressures for special education placement from parents and regular educators. Although the evidence is by no means definitive, there is increasing reason to believe that LD is "a middle class disease" and that more often than not parents actively seek the label of LD to obtain resource room help for their children. As cited previously, Warner et al. (1980) could find very few differences between low achievers and LD pupils. In Schumaker et al. (1980), however, the same team of researchers found that the two groups of students did differ in the degree of support they received from their parents. "LD pupils" tended to have more supportive parents; the researchers conjectured that parent intervention might explain why some low achievers were identified as LD and others were not:

> Because of their tendency to be supportive and go to the school at signs of trouble, these parents may have sought the extra help they perceived their children to need and, through this advocacy, may have caused their children to be labeled learning disabled (p. 18).

In interviews with a representative sample of special education directors, Shepard and Smith (1981) were told that parents who took an active part in

the staffing of their child were usually pushing for special education placement. Instances in which parents were active but were against identification were reported to be very rare. In case study research aimed at understanding the staffing process, Smith (1982) described the case of a wealthy mother who persisted in her demand for an LD diagnosis for her son, eliciting the help of an outside consultant who attended the staffing conference with her. Neither the school nor the child himself thought he was handicapped. The first assessment and staffing did not result in placement. But the next year the mother again made a referral, the assessment process was repeated, and the child was eventually placed for resource room help. Smith noted that the hint of litigation may have influenced the clinicians' decision.

Parental attitudes toward the LD label for their children in no way resembles the social stigma associated with mental retardation even a generation ago. Although there are occasionally parents, especially from minority groups, who resist calling their children "handicapped," such reluctance is relatively rare. Instead, professionals report that parents are often eager and enthusiastic about the diagnosis. For some, the label is as good as an explanation for poor achievement and brings a sense of relief; Smith (1982) hypothesized that at least for one mother the label might be important for shifting the locus of blame for her son's problems from the home to the school. Surely, there are psychological factors that contribute to the attractiveness of the label that we understand very poorly.

Regular educators also exert strong pressure on special educators for help in dealing with problem children. An inconsistency occurs between how principals act as a group (e.g., they might complain collectively about how

big a bite special education takes from the general funds) and how they act individually. On a case-by-case basis, the school principal most often acts to support the regular classroom teacher in making a request for special education services. Although regular classroom teachers and principals may object in the abstract to an expanding special education population and budget, in individual cases they lobby for placement. In fact, they sometimes even argue that given the resources of special education, help with a marginal child is their due. They apply pressure in individual cases saying, "You owe it to us" or "Now it's our turn."

Teaching and System Failures

Shepard and Smith (1981) had suggested both an admirable and reprehensible explanation for overidentification in the LD category. In addition to the commendable motive of helping children with severe needs, a less commendable motive can also be described for misidentification, namely, removing troublesome and hard-to-teach children from the regular classroom. In the Colorado study, some of the LD cases who did not have any of the indicators of LD and did not qualify for other handicapped subgroups were actually above grade level on nationally normed tests. Some of these had minor behavior problems as their only abnormal charcteristics. Some had complete files but not a single indicator of LD or other learning or behavior problem. One director mused, in fact, that there were certain "chronic referring teachers" who refused to deal with any heterogeneity in their classrooms. As soon as the lowest student was referred and placed, the next lowest child would be a candidate for special education.

Coles (1978) proposed a radical thesis that labeling a child learning disabled is a way of blaming children for what is actually the failure of

schools to provide adequate education for all. For the 20% to 25% of LD cases who have no signs of a handicap or who are not seriously below grade level, it is more reasonable to propose that the disorder is in the school environment rather than in the child. However, in a qualitative analysis of 200 representative cases Shepard and Smith (1981) found that "teaching problems" were mentioned by specialists in less than 1% of the cases as a possible source of the problem. We acknowledged that teaching or situational adaptations may have been considered in those cases that had been referred but not placed in LD. However, given the extent of overidentification, it does not appear that the question of problems in the school setting is raised often enough.

Several factors may predispose specialists to overlook teaching failures and help label a normal child LD. In the technical section of the paper, the problem of referral bias is discussed again. There the issue is the extent to which early labeling of the problem is merely confirmed by final labeling. A prior issue exists, however, in the extent to which specialists believe that their role (after a referral) is to name <u>the problem</u> rather than evaluate <u>whether</u> a problem exists. Learning disabilities teachers especially may perceive that they have low status in a school and may need to prove their worth by confirming low-scoring normal children as handicapped rather than confronting the classroom teachers with suggestions for modifying their teaching strategies.

The Consequences of Overidentification

In the first major section of the paper, evidence has been presented to support the claim that many nonhandicapped children are being improperly diagnosed as LD. Several social factors were identified that interact with

definitional ambiguity and assessment problems to promote overidentification. In the remaining technical portion of the paper assessment practices that contribute to misidentification or to systematic overidentification are explained. Thoughout the report there is the implication or assumption that overidentification is bad and should be avoided through the use of more correct assessment and better awareness of the social pressures. In this section, the positive and negative consequences of overidentification are enumerated to explain why overidentification is considered to be more bad than good.

The most important positive consequence of being identified LD (when you are not) should be fairly obvious--you get extra help. On the average, "mild" LD cases recieve an hour of resource room help per day where they are taught one-on-one by a specially trained teacher or participate in very small groups. According to our survey data (Shepard & Smith, 1981), about half of the resource room instructional time is spent on remedial tutoring, which pupils who are behind in school need whether they are LD or not. (Other major blocks of time are devoted to remediation of underlying processing disorders or to modification of inappropriate behaviors and effect.)

Other positive benefits of the LD label include more elusive psychological gains for parents who are glad to have the extra help and relieved to have a socially desirable explanation for their child's slow school achievement. In some states with minimum competency tests for high school graduates, students with a handicapped label may be excused from the test but still receive a marketable diploma.

The negative consequences of misidentifying a child LD include the potentially harmful effects of the label itself. The diagnosis clearly says

that the problem is in the child, although not his or her fault. The label could have a detrimental effect on self-confidence and could subtly lower expectations. MacMillan and Meyers (1979) have reviewed extensively the research on educational labeling of handicapped learners and concluded that the presumed negative effects have not been demonstrated empirically; moreover, the effects are likely to be so complex that they cannot be uncovered by ordinary research controls. Nevertheless, even without the empirical proof, one has to be concerned that there may be some instances where the label could be harmful especially for members of minority groups. The problem is potentially more serious because the issues of labeling are often dismissed for the LD category because the majority of parents and educators believe it is a "nonstigmatizing" label.

Low achievers who are labeled LD also are affected negatively if the special help they receive is inappropriate. LD-identified children leave their regular classrooms for an hour every day and miss regular instruction. If the resource room help is not clearly superior to regular instruction, they will lose ground. LD "treatments" intended to fix underlying processing deficits are of questionable validity even for correctly identified LD cases (Arter & Jenkins, 1979), and they are surely inappropriate for slow learners and bilingual children who have been mislabeled LD.

In addition to the potential negative consequences of overidentification for the individual child, there are also negative consequences for the educational system. The dollar costs of identification and due-process procedures are excessive and unnecessary for the nonhandicapped child who needs remedial help. Shepard and Smith (1981) found that almost half of the special education funds available for the LD

category (combination of federal, state, and local dollars) went into the assessment and staffing process each year. This finding argues against the positive motive for calling a child LD to get special help since almost half of the special resource is siphoned off to support the costs of identification. Craig, Myers, and Wujek (1982) reached similar conclusions in their discussion of policy implications of the California LD study: "Certainly, for each student that can be adequately served in a program other than special education, the costs of comprehensive assessment and due process procedures required by special education regulations can be avoided" (p. XI).

A less tangible negative effect of overidentification is that debilitating effect on classroom teachers (Beery, undated). By referring all learning problems out of the regular classroom, the teacher becomes less and less able to deal with a variety of learning needs. Normal variations in learner abilities then begin to look abnormal to teachers with a narrow range of instructional strategies.

The excessive costs of identification (which take away half of the extra resource gained), the potential harm of labeling, and the inappropriateness of treatments are negative effects of overidentification that outweigh the good effect of remedial help. Remedial programs that do not require a handicapped label and a wider repertoire of skills for regular classroom teachers would be preferred ways to achieve the good ends without the negative consequences.

Technical Issues in LD Assessment

This section of the paper focuses on psychometric and technical problems in LD diagnosis. How should tests and behavioral data be used to determine that a child is LD? Here the assumption is made that the reader or specialist has a scientific purpose in mind when making the assessment; i.e., "Is the child validly LD?" not, "What is the most appropriate placement, given the child's needs?" The purpose of the entire first portion of the paper was to try and separate what Hewett and Forness (1974) called the service and science motivations of the profession. There is no point in trying to be accurate in assessing a difficult and elusive construct if the institutional and social pressures supersede the data.

This section is also organized by subheadings that correspond to different steps in the assessment process, to operational measures of the definitional elements, or to general conceptual issues. On some topics, such as the adequacy of standardized tests, there is an enormous body of previous writing. For these topics a brief summary is provided with reference to more in-depth presentations. In contrast, some conceptual points are developed here that have not been made elsewhere. These ideas represent my attempt to make generalizations about problem areas and to suggest solutions based on extensive study of the LD identification process.

Referral Bias

When classroom teachers refer a child for special education help, they have already reached a decision in their own minds that, to some degree, the child has

a problem that is more serious than they can deal with, given the demands of a full classroom of children, etc. They may tacitly or explicitly ask specialist to sanction this conclusion. This expectation creates the social or collegial pressure for placement discussed in the first part of the paper. Separate from the social demand, however, there may be a cognitive bias associated with the referral that exists even if the classroom teacher were completely neutral about the necessity for placement. Given the ambiguity of criteria for a subjectively determined disorder, there is latitude for clinicians to see in a child what they are predisposed to see. Moreover, there is some research evidence to suggest that the naming of "the problem" in the referral creates such a predisposition. For example, special education trainees who were told they would be observing an emotionally disturbed child rated the behaviors of a normal child much more negatively than those who were told to expect a normal child (Foster, Ysseldyke, & Reese, 1975). In a study of simulated decision making, Ysseldyke and Algozzine (1981) found that the nature of the referral problem was more influential than sex, socioeconomic status, or attractiveness in the diagnostic decision, and that suggestion of an "academic" problem led to the judged likelihood of LD in a normal case (though the effect was not statistically significant). The effect of hypothetical behavior problems was much more strongly influential, leading to the diagnosis of emotionally disturbed (ED) in a normal case. In Ysseldyke et al. (1981), decision makers reported that the reason for referral had a pronounced effect on outcome decisions. Ysseldyke et al. also noted that 95% of all referrals in New York City result in placement in special education. More research on referrals and intervention prior to referral are needed so that the results of assessments will not be determined before they begin.

Normal Variability and Clinicians' Vertigo

Based on my reading of hundreds of individual specialists' reports in a representative sample of LD pupil files, I have developed the following hypothesis about why professionals sometimes see LD when it is not there. I believe that referral bias (i.e., the teacher sounds an alarm that is confirmed by the clinician) combines with misinformation about normal variability to cause professionals to interpret behavior that is within the normal range as if it were abnormal. In other words, clinicians develop a kind of "vertigo." Just as a pilot quickly loses his bearings when the horizon is obscured by clouds, so the specialist who sees only "referred cases" day after day loses track of what constitutes normal performance.

Measurement specialists have had considerable experience indicating that regular classroom teachers, at least, have difficulty internalizing relevant normative comparisons. Regular teachers are very good at ranking the relative achievement of the students within their classrooms (r with standardized tests = .53 to .87, Kretke et al., 1976). But they are very poor at comparing the standing of their classrooms to national averages. Instead, all teachers tend to think that their classes are "average." This tendency to adopt relativistic norms causes serious problems occasionally, as when an average-ability student (IQ = 100) in a high-socioeconomic, high-achieving school district is counseled not to apply for college since he or she is not likely to be successful.

In the diagnosis of LD there is evidence that specialists tend to interpret certain signs or test-score patterns as if they were abnormal when they are actually quite normal (i.e., the same pattern occurs for 25% or 35% of the normal population). To this effect, specific research findings by Kaufman (1976a, 1976b) will be described in later sections on significant discrepancies and subtest scatter.

As an example, however, note that clinicians believe that the average difference between verbal and performance IQs on the WISC-R is 4 to 6 points, whereas the actual average difference is 10 points. Similarly, clinicians consider a verbal-performance discrepancy of 12 or 15 points to be extreme and unusual, yet scores this different were observed for 33% and 25% of the standardization sample respectively (Kaufman, 1976b).

Clinicians who see only "at-risk" or referred children lose their bearings and do not have accurate internalized norms of typical variability. Furthermore, as we will see in a later section on clinicians' technical knowledge, specialists may not have adequate technical expertise to use normative data to correct their misperceptions. Part of my reading of LD case files involved comparing actual test scores with the interpretive sentences written about them. It appears that, by and large, specialists have adequate preparation to interpret measures of central tendency (mean, median), but frequently do not know how to use indices of variability (e.g., standard deviation).

Let me offer the following example to illustrate why variance as well as central tendency is important for understanding normalness. Suppose the mother of a three-year old girl was told that her daughter's weight (26.5 lbs.) was the same as an average two-year-old's. This sounds like a serious deficiency, but is it? In fact, compared with other three-year-old girls she is at the 10th percentile, small but not abnormal. If, however, the same three-year-old girl were the same <u>height</u> as an average two-year-old (34"), she would be smaller than .5% (one-half of one percent) of all three-year-old girls, perhaps cause for concern. Being "a year behind" on the two distributions (height and weight) has quite different meanings because the variances are different. Because the variances are different, the overlap between the two age distributions is different.

Clinicians appear to ignore normal variability when they expect everyone to be at the mean for their age group and treat any short fall as evidence of a serious problem or abnormality. In fact, very few children of a given age score exactly at the mean on any measure. A standard deviation has to be added on either side of the mean to include a majority (two-thirds) of the age group. Further, it very much depends on the particular academic test or developmental scale whether the standard deviation spans years or only a few months. Therefore, unless the particular measure is known, it is not possible to know whether being "a year behind" is normal or abnormal. It has been my observation that even experienced clinicians are inclined to overinterpret below-age-level performance, especially in young children. For example, a screening device shows a four-year-old's language development to be at the "three-year-old level." The mother and therapist are alarmed and seek intensive treatment. But just as in the height-weight example, we have to ask how rare and hence how deficient this performance is. Another way to ask the same question is: "How much do the two distributions overlap?" Is the median score for three-year-olds at the 2nd or 25th percentile of four-year-olds?

In the diagnosis of LD, patterns of scores are also important as well as single scores. But patterns of scores can likewise be rare or frequent. In order for a particular pattern to be taken as a sign that a child "is not functioning normally," it has to be relatively rare.

As a statistician, I would like to suggest that the so-called "average child" has become just as reified a term as some special education labels. A more appropriate conception would acknowledge that there are many ways to be normal. Surely all 80% or 90% of the children in the middle of the distribution should be considered normal. Therefore, a child should not be labeled as

disordered or abnormal unless his or her performance is so unusual that it falls outside of this range. Some would argue that all the children who are below the mean (50% in a normal distribution) have some developmental or learning "disability." My reasoning is quite different. I believe that our conceptions of what constitutes a disability is always relative to what others can do; i.e., a person who cannot fly is not disabled because it is not normal to fly. But, an eight-year-old child who cannot read at all after two years of instruction appears to be extremely unusual.

In later sections, specific technical problems are treated regarding the interpretation of discrepancy scores and test profiles as evidence of LD. From the foregoing argument it should be clear that these issues are not merely statistical. The question is not just what constitutes a reliable (i.e., statistically significant) pattern, but what patterns have validity as symptoms of a disability. A general, conceptual presentation about the problem of "normal variability" has been offered first so that the technical discussions will not be seen as an attempt to reduce diagnosis to simplistic formulae. Rather, the statistical computations and norms can act as guidelines to correct the "vertigo" or misperceptions that occur when we have poor experience with the range of normal variability.

Technical Adequacy of Tests

Standardized tests play a major role in the identification of children with learning disabilities.[2] For example, in their national survey of federally funded Child Service Demonstration Centers, Thurlow and Ysseldyke (1979) found that norm-referenced tests were used more often than any other source of information in making screening, classification, and placement decisions. (See also Poland et

[2] As one reviewer has noted, in the future, standardized tests may play relatively less of a role in the identification of LD. Because of recognized inadequacies in existing tests, many states are moving toward "non-test based approaches in the evaluation of LD." Behavioral data and clinical assessments are treated in a later section.

al., 1979; Thurlow, 1980). Indeed, given the earlier argument about the need for empirical data to establish what is normal, LD assessment procedures require a certain degree of standardization at least to ensure normative comparisons and validity evidence.

Unfortunately, psychometric inadequacies of tests used to assess LD are widely documented (Salvia & Ysseldyke, 1978; Ysseldyke, 1979). Coles (1978) reviewed the literature on the 10 measures most popularly used in the diagnosis of LD and concluded that none could validily distinguish LD from normal learners. In both Thurlow and Ysseldyke (1979) and Ysseldyke et al. (1980b), lists are provided of frequently used tests along with ratings of the adequacy of normative data and both reliabilaity and validity evidence. The criteria for evaluating these three test properties are derived from the APA test standards (1974); e.g., normative data must be from representative populations, reliability coefficients must be .90 or greater for individual decisions, and empirical evidence of validity must be provided for the particular use for which the test is intended. More than half of the tests reviewed failed on all three dimensions. Shepard and Smith (1981) integrated the ratings by Thurlow and Ysseldyke (1979) with findings from empirical studies (Arter and Jenkins, 1979) and individual test reviews found in Buros' Mental Measurements Yearbook and various professional journals. We concluded that of the 19 tests most frequently used in the identification of LD, only 5 met minimum standards for technical adequacy.

To understand what solutions exist to improve selection of technically adequate instruments, one must distinguish the nature of the problem for different categories of tests. As noted by Shepard and Smith (1981), there are good measures available for the assessment of IQ and achievement. But there are no valid and reliable instruments for measuring underlying processing disorders. As

will be discussed in the next section, the impediment to accurate assessment of IQ and achievement is not that adequate measures do not exist, but that professionals are so poorly informed that they choose bad tests instead of good.

The cumulative negative evidence against the validity of perceptual-processing tests has been reviewed by Arter and Jenkins (1979), Larsen and Hammill (1975), and Newcomer and Hammill (1975). Measures such as the ITPA lack sufficient subtest reliability to support the profile analyses for which they were intended (Lumsden, 1978). Further, they do not have discriminant validity from IQ (i.e., they are redundant with IQ), even though the inferences made suppose that the cognitive process measured is separable from reasoning ability (Larsen, Rogers, & Sowell, 1976). Both the theory and measurement of underlying processes have proved so unsatisfactory that the attribution of LD to a dysfunction in the "basic psychological processes" has been omitted from the new NJCLD definition (Hammill et al., 1981). Similarly, Harber (1981) suggested that attention has shifted away from trying to measure psychological process toward appropriate assessment of discrepancies in achievement areas. Of course, in the absence of measures which can legitimately connect a learning problem to an internal disorder, more reliance must be placed on the other means for inferring a specific intrinsic dysfunction, i.e., evidence of discrepancy and exclusion criteria. Both are addressed in later sections.

Determining that a child is LD requires accurate assessment of general intellectual functioning. For this purpose, the WISC-R is clearly the superior measure. Reliabilities are on the order of .95 and construct validity evidence is contributed by more than 1,000 research studies. For preschool or adult populations the WPPSI, Stanford Binet, or WAIS may be preferable. For children who are not from the majority culture, nonverbal measures such as the Ravens' or

supplemental measures of adaptive behavior may be preferre[illegible] his latter topic is covered in the section on bias in assessment.

Other measures of "IQ" are sometimes used in place of the WISC-R or in addition to it. The Detroit does not have adequate subtest reliabilities and has no evidence of validity. The Slosson has unknown reliability and was normed only on a clinical sample of retardates. (See more detailed reviews of tests in Shepard and Smith, 1981.) The PPVT is so clearly a vocabulary test rather than a measure of intelligence that advertisements for the new version call it a test of "hearing vocabulary." Nevertheless, some specialists continue to use it as the only indicator of IQ. Reasons for using inadequate tests instead of the WISC-R include overloading of psychologists' time, especially in rural districts, and, as we will see in the next section, misinformation on the part of specialists about the adequacy of tests.

Achievement in particular school subjects is probably the area in all of education and psychology that can be measures with the greatest validity. Unlike measurement of intelligence, the inferences required to connect test-taking behavior to an underlying construct are not so strong. The intended content domain of the test can be specified with much greater accuracy and concreteness. Numerous excellent test batteries have been developed, with substantial empirical documentation to measure achievement in basic skill areas such as reading, mathematics, language, and spelling. Ironically, the most carefully developed achievement measures tend to be group administered tests. This is ironic because many specialists have been trained to believe that individually administered tests are always better (as is the case with the WISC-R compared with group IQ tests). Excellent group achievement tests include the Comprehensive Test of Basic Skills and Stanford Achievement Tests. Of course, in individual cases specialists may

conclude that the presenting difficulty of a pupil is such that a standard paper-and-pencil testing situation is inappropriate. In these cases one of the technically adequate individual measures should be used, such as the PIAT or Woodcock Reading Mastery test.

Details regarding the deficiences of other popular achievement measures, including the WRAT and Key Math test, are given in Shepard and Smith (1981). Two general conceptual points can be made about their inadequacies. First, as the name wide-range implies, the WRAT spans an enormous range of curricular content. It would be a good "quick and dirty" screening device for locating the general grade-level placement of a student moving to a district from out of state. But, because it covers such a wide range, there are relatively few items at any given level and hence less accurate assessment. The WRAT is not recommended for indivudual decisions as important as special education placement. In contrast, the Key Math test has much greater content validity but absolutely no normative data. In the manual means are given for each grade placement, but not standard deviations. As was explained in the section on "normal variability," clinicians are then often misled because they think any score falling in the grade below the child's current grade placement is seriously deficient (but early in the school year this could be true for 50% of the child's classmates). For this reason Shepard and Smith (1981) gave the Key Math test a grade of "C" for diagnostic purposes but an "A" for instructional planning uses where normative referents are not crucial.

Specialists' Knowledge of Test Adequacy and Measurement Concepts

Early in this report it was suggested that many factors contribute to overidentification of LD, including charitable motives to help children with special

needs. Although it would be naive to think that solving the technical problems associated with LD diagnosis would be sufficient in face of the complex social pressures, nevertheless, one of the serious problems contributing to overidentification of LD is the lack of adequate technical knowledge on the part of specialists. The focus of this paper is on technical problems associated with LD identification and on the types of specialist training and insights that would lead to improved diagnosis.

The problem of psychometrically inadequate tests, discussed above, is compounded because substantial numbers of specialists do not know the difference between good and bad tests. Ysseldyke et al. (1980b) conducted a simulation study of special placement decision making. They found that school personnel initially chose inadequate tests as often as adequate ones and that the more tests they chose, the more inadequate ones they included. This often cited study by Ysseldyke et al. (1980b) was criticized by Wright (1980) on several grounds; e.g., the available pool of instruments included more inadequate than adequate tests, professionals who do not usually give tests participated as well as specialists, and the case presented was bogus since the data were all within the normal range. In keeping with the earlier hypotheses about referral bias and clinicians' vertigo, it is interesting to note that when presented with normal data in this study, clinicians tended to keep testing (with inadequate measures) rather than stop and conclude that the child was normal. This point was made by Ysseldyke et al. (1980a) in their rebuttal of Wright (1980). Other criticisms suggested by Wright have been refuted by subsequent studies. For example, La Grow and Prochnow-La Grow (1982) surveyed only school psychologists in actual testing practice where the choice of instruments was not constrained. They found that only two of the most widely used tests met minimum technical standards.

Shepard and Smith (1981) addressed directly the quetion of whether specialists knew when they were using bad tests, by asking psychologists, LD teachers, and speech-language specialists to rate the reliability and validity of tests for LD diagnosis. Findings from these surveys are also reported in Davis and Shepard (in press). The results indicated that from one-third to one-half of the professionals were misinformed about the technical properties of tests they used often. For example, 46% of the LD teachers and 55% of the psychologists rated the Beery Developmental Test of Visual-Motor Integration (VMI) as "having adequate research evidence for its validity in diagnosing LD" even though no empirical evidence has been published to support the test. Speech-language specialists rated the Detroit and WISC-R as equally valid measures of IQ.

Shepard and Smith (1981) also found that specialists often selected technically inadequate measures even when more valid instruments were available. Later we concluded (Shepard and Smith, in press) that these choices tended to follow traditional habits associated with each professional group:

> For example, the Peabody Picture Vocabulary Test was the IQ measure used most frequently by LD teachers and speech/language specialists, followed by the Detroit. (The practice of each professional giving their own IQ test explains why 25% of the LD cases had had three or more IQ tests as part of their initial assessment). The WRAT was still the favorite achievement measure of school psychologists (40% of that group said it had adequate validity for LD diagnosis) (p. 00).

Perhaps even more serious than inadequate tests and lack of specialist knowledge about test adequacy is the concomitant lack of adequate preparation in test-score interpretation. McDaniels (1979), for example, identified personnel training as a more serious need than development of new measurement technology. In this paper, specific misconceptions on the part of specialists are dealt with in

appropriate sections pertaining to technical points such as significant discrepancy and subtest scatter. A more general review of professional competence regarding assessment is provided by Bennett (1981). He cited both opinion data and research findings to support the conclusion that professionals may lack basic test-score interpretation skills.

In a study focused specifically on LD specialists, Bennett and Shepard (1982) found that the LD professionals could answer only half of the questions on an introductory measurement course examination. Questions dealt with basic concepts such as interpreting relative performance on two tests with different scales and using standard errors of measurement to establish confidence intervals around observed scores. Although Bennett and Sheperd cautioned that their findings may not be indicative of actual assessment practice, the analyses of real cases conducted by Shepard and Smith (1981) revealed the specific links whereby inadequate technical knowledge led to inappropriate score interpretation which led in turn to invalid identification. Davis and Shepard (in press) reported, for example, that half of the LD teachers were unable to identify a significant discrepancy in a fairly simple hypothetical problem. This finding was corroborated by analyses of individual specialists' reports in real LD case files. A common mistake, for example, was for specialists to treat an IQ of 90 as if it were at the median (50th percentile) instead of the 25th percentile since it is "within the normal range." As suggested in an earlier section, specialists may not have even minimal statistical and measurement competencies which are needed if professionals are to keep their bearings in distinguishing normal and abnormal performance. Specific examples are explained in the following sections.

Significant Ability-Achievement Discrepancy

In an earlier section, the concept of discrepancy, between intellectual ability and actual achievement, was presenteda as a key element in the construct of LD. Although discrepancy is not mentioned in either the old or the new definitions, it is the primary, concrete criterion in the guidelines that accompany the federal definition (USOE, 1977, p. 65083). Conceptually the notion of discrepancy or anomalous intellectual functioning is essential to the meaning of LD, since a learning disability involves intrinsic intellectual functioning but is clearly distinct from mental retardation. LD is distinguished from mental retardation by the feature of discrepancy; i.e., in LD the learning difficulty is unusual whereas with retardation the dysfunction is consistent across many areas of cognitive functioning.

Given that ability-achievement discrepancy is essential to the LD construct, it follows that assessment of LD should involve comparing measures of both intellectual ability and achievement. Numerous formulae exist, in fact, for computing the significance of discrepancy between two observed test scores. It should be made emphatically clear, however, that the purpose of these formulae is to serve as a guideline for interpreting the magnitude of the differences, not to stand as the sole criterion for LD diagnosis. As concluded by Shepard and Smith (1981):

> Rules and criteria can be improved. They cannot, however, force valid placements. As with many psychological constructs, the validity of LD identification cannot be reduced to simplistic statistical rules. Minimal criteria for the reliability and discriminant validity of both formal and informal assessments can be established, but ultimately the integration of separate pieces of diagnostic information must rest on professional judgment.

Detailed reviews of the strengths and weaknesses of various discrepancy formulae have been provided by Cone and Wilson (1981), McLeod (1979), and Shepard (1980). Some of the most widely used procedures are wrong on both technical and conceptual grounds. For example, simply computing a child's years below grade level ignores both the child's IQ and the fact that natural variability increases with grade level (so being a year behind is much more serious at grade 2 than grade 9). Formulae that were invented to try and take ability into account, e.g., Harris (1970, Bond and Tinker (1967), and the proposed federal formula (BEH 1976), were psychometrically inadequate because they misrepresented the empirical relationship between IQ and achievement. (Basically, the authors made the mistake of assuming that what was true at the mean was true elsewhere in the distribution.) Erickson (1975) demonstrated that these formulae, like the below-grade-level criterion, would identify slow learners and all low achievers rather than the learning disabled.

More complex and psychometrically sound procedures for quantifying discrepancy based on regression analysis have been advanced by McLeod (1979) and Shepard (1980). See Cone and Wilson (1981) for a detailed explanation as to why these methods are to be preferred. Basically, regression analysis takes into account not only the random error associated with the measurement of both IQ and achievement but also the regression to the mean that will occur because of the imperfect correlation between IQ and achievement.

If one does not have access to co-normed tests for which regression equations are known (Shepard, 1980), the formula for computing the standard error of the difference is the next best method for evaluating discrepancy (Salvia & Ysseldyke, 1978; Thorndike & Hagen, 1977; see Appendix A for computational

examples). Keep in mind that the purpose of these formulae is to estimate how big a difference could be expected just by chance, given the measurement error in each of the tests. Thus, only IQ-achievement differences that are greater than two standard errors of the difference should be treated as reliable differences.

The formula for computing the standard error of the difference can even be used with estimated values for the between-test correlations, since we presume that a "significant" difference or lack thereof is not going to be interpreted rigidly. Rather, given that specialists tend to overinterpret small differences, these computations serve as a safeguard, giving us a rough conceptual minimum when interpreting differences. It should be pointed out that once the standard error of the difference has been computed for a given pair of tests, say the WISC-R and PIAT, it can serve as a constant yardstick for evaluating discrepancy scores every time that pair of tests is administered. Therefore, the procedures can become quite easy to use with familiar pairs of tests so long as specialists know how to convert raw scores to standard scores. (See Appendix A standard erors formulae, for tables of standard errors for the most frequently used test pairs, and for computational examples.)

Cone and Wilson (1981) note that sometimes so little is known about either the reliability of tests used or the correlation between tests that one has to resort to simple comparisons of standard scores (Erickson, 1975; Hanna, Dyck, & Holen, 1979). In other words, is the child's percentile rank on the achievement test roughly the same as his or her percentile rank on the IQ test? Although the psychometric errors in this approach are well documented (Cone & Wilson, 1981; Shepard, 1980), I agree with Cone and Wilson that standard score comparisons would represent a substantial improvement over current practices. In the Davis

and Shepard (in press) study cited earlier, half of the LD teachers and school psychologists could not identify a significant discrepancy correctly. In the multiple-choice question posed, the answer would have been trivially easy if respondents had known that an IQ of 90 was at the 25th percentile. Because they were not accustomed to making this conversion to percentiles or common standard scores, many though that reading achievement at the 35th or 28th percentile was significantly below an IQ of 90 expectancy, when it is in fact above it.

Use of either z score comparisons or standard error of the difference computations would help specialists avoid certain common errors that currently contribute to overidentification. For example, in reading individual specialists' reports, Shepard and Smith (1981) noted that a frequent practice was to treat an IQ of 90 as if it were "in the average range" and therefore to expect achievement to be at the 50th percentile. Not only is this expectation considerably in error since 90 is at the 25th percentile, but by this practice clinicians are implicitly assuming an asymmetrical confidence interval around the observed score. That is, allowing for measurement error, they believe the true score could be 95 but not 85. It is at this stage of interpretation that clinicians may be inclined to slant the meaning of the assessment data so that a marginal case can receive services. Similar errors are made whenever specialists interpret below-grade-level scores (e.g., a fourth grader scoring at 3.2) as signs of serious deficiency without realizing that a large percentage of fourth graders may have similar scores and that the percentile rank may be consistent with the child's IQ.

Although discrepancy computations can help to establish minimum reliable differences, two cautions are offered as to why reliable discrepancy is not automatically synonymous with LD. First, it is widely recognized that a learning

disability could depress the observed IQ measure and hence prevent a discrepancy (Danielson & Bauer, 1978). This caution does not mean, however, that every low IQ score should be dismissed as invalid. Again, in current practice, this explanation for lack of discrepancy is invoked much more frequently than it could possibly be true, thus placing many slow learners in LD. Before claiming that measured IQ is depressed, clinicians should have some other indication of higher intellectual functioning such as extreme verbal-performance discrepancy and average achievement in math but deficient reading.

Significant discrepancies can also be signs of poor motivation, absence from school, or normal variation. Reliability is necessary but not sufficient for validity. Just as Kaufman (1976b) found that reliable verbal=performance discrepancies are not rare (occurring for one-third of the normal population), so significant IQ-achievement discrepancies will occur for many individuals who are not LD. It is especially important that specialists not try to find a problem by continuing to test until a discrepancy occurs. The more tests that are given, of course, the greater the probability of finding a significant difference just by chance.

Interpreting Subtest Scatter

Because learning disabilities are believed to be specific disorders in an otherwise able child, specialists will often look for perturbations in test performance as a sign of LD. When a child exhibits very different abilities on different types of tasks within a test, the subtest scores are said to have significant "scatter." If a child's level of performance is uniform across various subtests, the result is called a "flat profile."

For subtest scatter to be a valid indicator of LD, at a minimum the apparent variability in abilities must be reliable (be greater than chance). If the child's strengths and weaknesses shifted from one testing to the next, it might suggest poor effort or attention during the tests but not an enduring pattern of inherent abilities and disabilities. Tests such as the ITPA and Detroit do not have adequate subtest reliabilities to support the types of profile interpretations usually made. Even on tests with generally better subtest reliabilities, such as the WISC-R, the amount of fluctuation required in the profile, before the differences could be considered reliable, is quite large. Salvia and Ysseldyke (1978) provided an example of a WISC-R profile that appears to be irregular but which only has one statistically reliable, deviant subtest score.

As was the case with significant discrepancy scores, reliability is necessary but not sufficient for validity. For scatter to have validity as an indicator of LD, it has to be consistently found in known LD children and not found in normal children. Salvia and Ysseldyke (1978, p. 410) cited this as the difficulty with trying to use scatter as a diagnostic tool; that is, it appears too often in normals. Although there may be a weak relationship between scatter and clinically identified groups, the relationship is not sufficient for making individual diagnoses. They quoted Cronbach (1960), "This type of analysis is no longer depended upon because empirical checks show that pattern analysis has little validity" (p. 192).

Clinicians who work only with "at-risk" children in the population may not have the opportunity to build up experience with the amount of scatter typically found in average and normal children. Kaufman (1976a, 1976b) used the standardization sample from the WISC-R to construct "norms" for interpreting

subtest differences. The results are surprising since the amount of difference that is "usual" seems counter-intuitive. Using a criterion of 15% in the standardization sample as a cutoff for abnormal occurrences, Kaufman (1976a) concluded that "a 10-test range of 6 to 15 or 3 to 12 would not be considered unusual" (p. 163). Clinicians frequently cite a range of this amount as evidence supporting a LD diagnosis, since this variation does meet requirements for reliability. However, if large ranges are normal, they cannot be valid signs of LD.

Readers who still harbor some intuitive, persistent faith that subtest profiles can yield valid diagnoses should consult the three-part series of articles in the September, October, and November 1981 issues of the Journal of Learning Disabilities. For example, Reynolds and Gutkin (1981) examined four different indices of scatter on the WPPSI and concluded that "what were previously believed to be unusual amounts of within test variability of performance for individual children were found to characterize the profiles of many normally functioning children" (p. 460). The Bannatyne (1968) method for recategorizing WISC-R subtests has been regarded as particularly promising because the patterns were empirically derived originally. Although this procedure may be an important research tool for understanding particular subtypes of LD, Henry and Wittman (1981) found that the Bannatyne patterns could not differentiate LD from normal students and "might even contribute to misdiagnosis" (p. 517). Kaufman (1981b) reviewed the research on WISC-R subtest and scatter interpretations and concluded that most clinical stereotypes do not hold true. Although there may still be legitimate, small subgroups of LD for whom extreme patterns are characteristic, there is not now sufficient evidence to support LD diagnosis on the basis of these profiles. As a rough rule of thumb clinicians should realize that to

call a V-P discrepancy abnormal would require a difference of 26 points; similarly, abnormal scatter would require a scaled score range of 12 points or more (Kaufman, 1981a). On a different test, the McCarthy Scales of Children's Abilities, Goh and Simons (1980) likewise found that LD and general education children had similar amounts of scatter.

Using Age Norms to Evaluate Processing Deficits

The serious deficiencies in the psychometric properties of tests used to measure underlying psychological processes have already been belabored. Sometimes clinicians continue to use these measures despite their unreliability and questionable validity saying, "It is the only thing available" or "I'm only using it clinically to explore hypotheses." When specialists persist in using processing tests, there is an additional interpretation problem that can lead to invalid diagnosis of LD. Clinicians frequently use age norms (i.e., the median performance level for children of a given age) to determine whether a child has a processing deficit. For example, in Colorado, for both standardized tests and informal assessments, significant processing deficits were defined by the following criteria:

Ages	Years of Deficit
3-8	1 year
9-12	1 1/2 years
13-21	2 years

This method of evaluating processing skills in relation to age-group medians is contradictory to the ability-achievement discrepancy component of the LD definition. Because intelligence is correlated with information processing abilities, it can be expected that children with low

intelligence and correspondingly poor achievement (i.e., no discrepancy) will also have low processing skills. Therefore, if low scores on processing tests are interpreted in relation to age norms rather than in relation to a child's own level of cognitive functioning, it is equivalent to defining LD as (severe) below average intelligence.

This criticism of the definition of processing deficits in relation to age medians does not imply that low IQ scores preclude interpretation of a processing disorder. Clinicians are faced with the problem that obtained IQ scores could be an underestimate of true ability if a processing problem interferes with test performance; this phenomenon would also prevent an ability-achievement discrepancy from being significant. But if this is the hypothesis to be tested, comparison with age norms does not help to resolve whether a child has low general intelligence which is also reflected on the processing test or a processing disorder which is depressing IQ test performance. The validity of the tests and the validity of the constructs they represent suggest the following approach: children with processing test scores at roughly the same level as their IQ scores (allowing for the unreliability in the tests) should not be identified as having a processing deficit unless there is consistent and statistically stable evidence of a processing dysfunction in a particular area that also coincides with the particular areas of poor performance on the IQ test. Furthermore, given the information in the preceding section regarding the amount of scatter that should be treated as normal, clinicians will have to develope more extreme criteria for interpreting symptoms of pathology. Recent evidence such as the Kaufman studies suggests that clinicians have been interpreting as abnormal patterns of scores and behaviors that are manifest by large segments of the normal population. It is "usual" for a

child with low intelligence to have low processing scores and to have considerable scatter within tests. Only a coherently interpretable picture of a particular processing problem should be allowed to refute the conclusion that the child has "normal" below average functioning.

Behavioral Indicators, Informal Assessment, and Clinical Hypothesis Testing

Three distinct activities that contribute to the assessment of LD are treated together because they are so entwined in current practice. Behavioral indicators are observable behaviors that constitute evidence of a disorder. Behaviors such as attention span may be assessed either by formal, standardized scales or by informal observations and checklists. Behavioral indicators and informal assessments are often thought to be synonymous because informal measures are most frequently used to assess behavior. Clinical hypothesis testing is more than a data gathering activity. Hypothesis testing is also a reasoning process whereby observed signs are tested logically for their fit or consistency with a presumed model of disorder.

Although the purpose of this report has been primarily to deal with the use of formal tests in LD assessment, a serious caveat should be issued regarding the use of behavioral indicators and informal assessments. A warning is especially in order, since informal observations are now seen as increasingly desirable precisely because standardized tests are inadequate. Some states and school districts are shifting to nontest criteria for identification of LD. There is the risk, however, that behavioral indicators and clinical observations will lead to just as many invalid placements as with test-based decisions. First, the folklore

and stereotypes regarding what behaviors are symptoms of LD are pervasive but largely untested. Many of the believed signs of LD can be traced back to Clements' (1966) list of symptoms gleaned from 100 studies of children with minimal brain dysfunction. The problem with that landmark survey is that the symptoms of MBD were taken at face value. That is, the signs were listed if they occurred in MBD samples. But, the discriminant validity of those symptoms for distinguishing MBD from normal children was never evaluated. This means that two symptoms of LD such as "short attention span" and "poor coordination" could both be significantly correlated with LD but could still be found (even in combination) in many more normal children than in LD children. In current research, certain signs such as attentional deficits appear to be very promising for understanding some subtypes of LD. At present, however, akin to the research on scatter, this characteristic probably accounts for only a very small subgroup of LD. Therefore, the symptom does not have diagnostic utility; i.e., for every correctly identified LD case with this symptom, 10 or 20 normal children would be found who also evidence the behavior.

A second serious problem exists for informal assessments whether of social behaviors or classroom achievement. The reliability and validity of these informal assessments is not known. Furthermore, they lack a normative basis for comparison. Extensive attention has been given in this report to the tendency for special education professionals to lose track of normal variability and hence see abnormality in every referred case. If special education professionals often forget that it is normal for many fourth graders to score at the third-grade level in math, even with norms tables and percentile ranks to remind them, how much more likely is it for diagnosticians to forget that perhaps 20% of fourth graders have

difficulty staying in their seats? Bennett (1982) offered specific suggestions for using informal assessments given that their technical adequacy cannot be known. My general advice would be (1) to use informal assessment more for instructional intervention than for diagnosis of a handicap and (2) to distrust one's conclusion of abnormality from classroom observations if there is not corresponding evidence on standardized measures.

Many specialists have not had adequate training in clinical judgment or hypothesis testing. Both survey results and case examples are presented in Davis and Shepard (1982, in press). In this particular regard school psychologists seem to be better trained than learning disabilities teachers and speech-language specialists. LD teachers especially equate clinical judgment with informal data collection and do not generally see the need for either consistency in observed signs or confirmation of diagnoses. In fact, one group of LD teachers believes it is contrary to the spirit of multidisciplinary team assessment to question the observations of others or to try to reconcile divergent findings. I have facetiously called this the "I'm OK, you're OK" model of clinical diagnosis. This attitude, reported in surveys of professionals (Davis & Shepard, in press), explains why in the study of representative LD pupil cases (Shepard & Smith, 1981), only 15% had highly consistent and coherent clinical signs of LD. Given that observational data can be very unreliable and that research evidence suggests that clinicians are inclined to see a problem when told to expect a problem (Foster, Ysseldyke, & Reese, 1975), assessment teams need to be much more active in challenging isolated and inconsistent evidence of LD. Since normal children also occasionally exhibit such patterns, the designation of LD should be reserved for only those cases with consistent evidence of the disability.

Exclusion and Bias

The purpose of the exclusionary clause in the definition of LD is to rule out other causes that are sufficient to account for the learning problem. If a child's performance is seriously depressed in all areas of cognitive functioning (both in and out of school), the more appropriate diagnosis is mental retardation. If a child with a hearing problem is behind in school, but improves with a hearing aid or a change in seating, the LD label is inappropriate.

Causes for exclusion may be in the child as in the above examples or in the child's environment. Cultural differences and insufficient opportunity to learn are examples of competing explanations for a child's lowered achievement which would argue against LD diagnosis.

The present overidentification of pupils in the LD category obviously suggests that the exclusionary rule is not being applied sufficiently in the determination of LD. Shepard and Smith (1981) found identifiable subgroups in the school LD population that included other handicaps such as educable mentally retarded, emotionally disturbed, and hearing handicapped. Also included were children with severe environmental problems (e.g., moving four times that school year or missing 30 days of school year after year) and children from non-English-speaking backgrounds. These apparent "diagnostic problems" are muddied, of course, by the professionals' motives and desire to provide help to students obviously in need.

The issue of exclusion for cultural differences is made more complicated by the correlation in the United States between ethnicity and economic status. Ethnic minorities are overrepresented in poorer classes, and poverty is known to have a negative relationship with school

achievement. It has already been pointed out that problems such as malnutrition associated with extreme poverty could have a debilitating and permanent effect on a child. Therefore, it is reasonable to expect a disproportionate number of LD cases from extremely poor families including ethnic minorities. But how big should this disproportion be, and how should this reasoning influence the assessment of an individual child?

Unfortunately, there is some evidence to suggest that clinicians adopt a blanket policy rather than trying to interpret the evidence in particular cases. In other words, they think that low SES should always argue for exclusion from LD or that low SES should always count toward the diagnosis of LD. Shepard and Smith (1981) found, for example, that there were two, almost equal, opposing groups of specialists who would, "other things being equal," consider linguistic differences as positive evidence for the determination of LD or against it. Overall the effect is still to include a substantial number of linguistically different children in the LD category (Shepard & Smith, 1981). In a simulation study of LD diagnosis, Frame et al. (1982) found that the low SES black case was classified as ineligible for special education more often than other SES and race categories. With real data, however, Tucker (1980) found substantial overrepresentation of blacks in the LD category. He attributed the burgeoning numbers in LD to concomitant social forces such as the civil rights movement and demands to protect black children from the stigma of EMR placement. Now, because LD is less stigmatizing and even popular, Tucker concludes that "LD can provide an excuse for a lower quality of schooling" (p. 105).

More appropriate application of the exclusion clause can be achieved by efforts at both the individual and aggregate levels. First, each school district should keep records of its own placement rates for

culturally different cases. The following are among the criteria suggested by Ysseldyke (1979) for evaluating compliance with protection in evaluation procedures, but staffing teams would do well to keep records and review their own performance in this regard.

> The LEA has a record of the number of children referred by individual teachers and regularly examines this record to ascertain the extent to which any one teacher has a history of over-referral of children from certain cultural groups or who demonstrate specific common characteristics.
>
> In all evaluation procedures, diagnostic personnel carefully consider the extent to which cultural differences or naturally occurring pupil characteristics may have biased the decision to refer a child.
>
> The LEA has established procedures for periodic evaluation of the extent to which cultural differences between teachers and children may lead to misinterpretation of child behavior and to unnecessary over-referral of children from specific cultural groups.
>
> The LEA regulary examines its referral patterns to ascertain the extent to which naturally occurring pupil characteristics affect the decision to refer children for consideration for special services (pp. 162-163).

In my experience school districts have avoided collecting data in this form unless they have been ordered to do so by the Office of Civil Rights. Perhaps there is a fear that disproportionate rates will automatically be misinterpreted by external parties. But, if staffing teams do not have these sorts of data about their own track records, it is not possible to determine whether there are any systematic tendencies ("biases") with marginal cases.

Appropriate implementation of the exclusion rule will also be improved if clinicians attempt to weigh the strength of evidence in each

individual case rather than imposing some general rule; e.g., all poor children are automatically excluded from LD. Just because there are compelling counter-examples to the preceding rule, however, does not mean that the opposite generalization is any more supportable; e.g., the child's mother is on welfare, so he must be LD. The only general rule should be to consider competing explanations for poor school performance in every case. Nationally, many poor children are achieving below-grade-level medians but within the normal range. If regular education has not been resourceful enough to meet their academic and emotional needs, it does not mean that all of these children are abnormal. To be called LD, a poor child should look deviant compared to this norm; he or she should have some other evidence of an intrinsic disorder rather than below-grade-level achievement.

Nondiscriminatory assessment and the exclusion rule are strongly linked. But nonbiased assessment is, of course, a much bigger issue, touching every aspect of data collection and data interpretation. For a more comprehensive treatment of issues and nonbiased assessment models than can be provided here, the reader is referred to Mercer (1979) and Ysseldyke (1979). Many of the problems are the same as I have already outlined; i.e., to what extent do the observed signs serve as valid indicators of the underlying construct? Are the tests used technically adequate? For minority children these problems are exacerbated because even measures which are technically adequate for use with children from the dominant culture may not be valid for some minority children. The two factors which are most likely to threaten the validity of routine assessment practices with minority children are differences in motivation (some minority children may not have a test-taking, task-oriented set) and

differences in exposure to relevant material. Many poor or minority children may not have learned the word gown and hence score lower on the vocabulary subtest of measured IQ. Obviously, for LD identification, the of ability or IQ. If ability is underestimated because of cultural bias in the IQ measure, the tendency would be to miss legitimate instances of LD because achievement would not be discrepant from IQ. Once again this problem cannot be solved by "blanket correction strategies." In reaction to the above problem clinicians sometimes assume that the child's IQ is 100 or in the normal range (90-110) and interpret all other data in this light. Since we have ample evidence to suggest that many slow learners are referred for assessment, this expectation is obviously too high (for blacks or whites), will create artificial discrepancies, and thereby will contribute to overidentification of LD.

To a large extent nondiscriminatory assessment must be an issue of personal values and social policy. As was said earlier, ambiguous definition and ambiguous symptoms make it possible for personal values and beliefs to influence the identification process whether consciously or unconsciously. Before the trend to overidentify linguistically different and black pupils (Tucker, 1980) can be reversed, clinicians will have to believe that false identifications can be harmful, especially for minority children. Currently, leanings are still in the opposite direction. The LD label and services are viewed so positively that below average but normal data are often construed as symptoms of LD to obtain services.

Summary and Recommendations

Problems in the assessment of LD were discussed in the context of social and institutional pressures that contribute to misidentification. Results from numerous large-scale studies all suggest that low achievers of many types are being overidentified in the LD category. One cause for misidentification, and especially systematic overidentification, is specialists' lack of technical knowledge. Particular technical errors or misconceptions were the focus of this report. Many other factors impinge, however, on the identification decision. Other causes of overidentification were reviewed, including ambiguity in the definition, the needs of nonhandicapped children for special services, parental demand, pressure from regular education, and the less admirable purpose of removing hard-to-teach children from the regular classroom.

Two general types of corrective changes were seen as necessary to forestall the nontechnical factors leading to of overidentification: (1) professionals will have to be convinced that the negative consequences of overidentification are serious, and (2) alternative programs will have to be provided for children who are not LD but are far behind in school. Although one-to-one instruction and special help are benefits of overidentification, harmful effects include labeling the child, instances of inappropriate services for non-LD pupils, the excessive costs of identification (nearly half of the special education resources available for the LD category), and the debilitating effects on regular education

teachers who learn to deal with a narrower and narrower range of learning abilities. Additional negative consequences (for which research evidence was not reviewed in this paper) include the confounding influences on LD research (the nature of the disorder cannot be studied in misidentified populations) and the potential political backlash against what appears to be a sham category.

Shepard and Smith (1981) found a moderately high correlation between district size and percent validity identified LD cases (valid identification was determined if cases met any one of eight different definitions of LD including statistical or clinical criteria). We conjectured that this relationship was due to two factors: (1) the very largest districts in the state were more closely scrutinized to keep their total numbers of LD small, and (2) larger districts tended to have more alternative programs such as Title I reading, bilingual education, and non-special-education resource rooms; therefore LD was more likely a placement of last resort in these districts. It should be comforting to those who believe that LD identification is a hopeless morass to see evidence that when a ceiling was placed on the number of placements, a greater percentage of valid identifications occurred; i.e., idependent researchers more often agreed with the staffing teams that the valid cases met either statistical or clinical criteria for LD.

Alternative programs such as bilingual education, intensive English instruction, or remedial reading tutoring have several advantages. First, the specific type of help needed can be obtained for much lower cost. If a handicapped label is not going to be affixed, most of the elaborate assessment and staffing costs, essential to ensure due process, are no longer necessary. Also, to the extent that special help can be provided in the context of the regular classroom, using a consultative teacher

model, the repertoire of the regular teacher would be increased rather than decreased.

Technical problems in the assessment of LD were reviewed in detail. The use of psychometrically inadequate tests and clinicians' lack of knowledge about test adequacy can create serious errors in the identificaiton of LD. Certain technical problems, such as the tendency for specialists to overinterpret small differences as if they were significant discrepancies, lead systematically to overidentification of LD. Furthermore, specialists continue to use stereotypical beliefs about LD characteristics such as subtest scatter and behavioral indicators; these presumed characteristics either have been disproven as in the case of subtest scatter or have had no empirical substantiation one way or the other as with behavioral indicators. Many signs now taken as evidence of LD do not have discriminant validity; i.e., they can't be used to differentiate LD from normal.

A major theme throughout this report was the need to recognize normal variability. It was argued that clinicians who see only referred children often develop a type of "vertigo" so they do not realize how similar many referred children are to others in the regular classroom. Furthermore, all of the research suggests that specialists tend to interpret as abnormal the discrepancies, scattered profiles, below-grade-level performance, and "inappropriate behaviors" that occur in large numbers of normal children. They expect all performance to be at the median with realizing that there is considerable spread around the median in the normal population. So that these misperceptions about what is normal might be overcome, this report stressed the need for normative comparisons for purposes of diagnosis. When the purpose of assessment is prescriptive,

i.e., to plan instructional interventions, normative data are not essential and criterion references measures may be preferred. Statistical essential and criterion references measures may be preferred. Statistical rules for interpreting significant ability-achievement discrepancies were reviewed, not because diagnosis can be reduced to simplistic formulae, but because these computations will give specialists a better insight into what constitutes a minimum reliable difference and also how this compares to valid rare discrepancies.

Many of the technical problems with LD assessment were seen to interact with the social and institutional pressures arguing for placement of low achievers. That is, ambiguous evidence is likely to be taken as evidence for LD because of these other pressures. Given the tendency to misjudge normal variability and the present substantial overidentification, however, the tendency should be in the other direction. If the data are weak or equivocal, the child should be called non-LD. A good motto might be "normal until proven otherwise."

References

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. Standards for educational and psychological tests (rev. ed.). Washington, D.C.: American Psychological Association, 1974.

Arter, J. A., & Jenkins, J. R. Differential diagnosis--prescriptive teaching: A critical appraisal. Review of Educational Research, 1979, 49, 517-555.

Bannatyne, A. Diagnosing learning disabilities and writing remedial prescriptions. Journal of Learning Disabilities, 1968, 1, 242-249.

Bateman, B. Learning disabilities--Yesterday, today and tomorrow. Exceptional Children, 1964, 31, 167-177.

Beery, K. E. Learning disabilities: The integrative functions. Unpublished paper, Institute for Independent Educational Research, undated (circa 1980).

Bennett, R. E. Professional competence and the assessment of exceptional children. Journal of Special Education, 1981, 15, 437-446.

Bennett, R. E. Cautions for the use of informal measures in the educational assessment of exceptional children. Journal of Learning Disabilities, 1982, 15, 337-339.

Bennett, F. E., & Shepherd, M. J. Basic measurement proficiency of learning disability specialists. Learning Disability Quarterly, 1982, 5, 177-184.

Bond, G., & Tinker, M. Reading difficulties: Their diagnosis and correction (2nd ed.). New York: Appleton-Century-Crofts, 1967.

Bureau of Education for the Handicapped. Federal Register, November 29, 1976, 41 (230), 52407.

Clements, S. D. Minimal brain dysfunction in children (NINDS Monograph No. 3, U. S. Public Health Service Publication No. 1415). Washington, D.C.: U.S. Government Printing Office, 1966.

Coles, G. S. The learning disabilities test battery: Empirical and social issues. Harvard Educational Review, 1978, 48, 313-340.

Cone, T. E., & Wilson, L. R. Quantifying a severe discrepancy: A critical analysis. Learning Disability Quarterly, 1981, 4, 359-371.

Craig, P. A.; Myers, E. G.; & Wujek, M. Independent evaluation of the California master plan for special education. Final Report. Menlo Park, California: SRI international, May 1982.

Cronbach, L. Essentials of psychological testing. New York: Harper, 1960.

Cronbach, L. J. Essentials of psychological testing (3rd ed.). New York: Harper & Row, 1970.

Cruickshank, W. M. Some issues facing the field of learning disability. Journal of Learning Disabilities, 1972, 5, 380-388.

Danielson, L. C., & Bauer, J. N. A formula-based classification of learning disabled children: An examination of the issues. Journal of Learning Disabilities, 1978, 11, 50-63.

Davis, W. A., & Shepard, L. A. The use of tests by LD teachers, school psychologists, and speech/language specialists in the diagnosis of learning disabilities. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, March 1982.

Davis, W. A., & Shepard, L. A. The use of tests and clinical judgment by specialists in the diagnosis of learning disabilities. Learning Disability Quarterly, in press.

Erickson, M. T. The z-score discrepancy method for identifying reading disabled children. Journal of Learning Disabilities, 1975, 8, 308-312.

Foster, G. G.; Ysseldyke, J. E.; & Reese, J. H. "I wouldn't have seen it if I hadn't believed it." Exceptional Children, 1975, 41, 469-473.

Frame, R. E.; Clarizio, H. F.; Porter, A. C.; & Vinsonhaler, J. R. Interclinician agreement and bias in school psychologists' diagnostic and treatment recommendations for a learning disabled child. Psychology in the Schools, 1982, 19, 319-327.

Gallagher, J. J. The sacred and profane uses of labeling. Mental Retardation, 1976, 14, 6-10.

Goh, D. S., & Simons, M. R. Comparison of learning disabled and general education children on the McCarthy Scales of Children's Abilities. Psychology in the Schools, 1980, 17, 429-436.

Hammill, D. D.; Leigh, J. E.; McNutt, G.; & Larsen, S. C. A new definition of learning disabilities. Learning Disability Quarterly, 1981, 4, 336-342.

Hanna, G. S.; Dyck, N. J.; & Holen, M. C. Objective analysis of achievement-aptitude discrepancies in LD classification. Learning Disability Quarterly, 1979, 2, 32-38.

Harber, J. R. Learning disability research: How far have we progressed? Learning Disability Quarterly, 1981, 4, 372-381.

Harris, A. How to increase reading abilities (5th ed.). New York: David McKay, 1970.

Henry, S. A., & Wittman, R. D. Diagnostic implications of Bannatyne's recategorized WISC-R scores for identifying learning disabled children. Journal of Learning Disabilities, 1981, 14, 517-520.

Hewett, F. M., & Forness, S. R. Education of exceptional learners. Boston: Forness, Allyn, & Bacon, 1974.

Kaufman, A. S. A new approach to the interpretation of test scatter on the WISC-R. Journal of Learning Disabilities, 1976, 9, 160-168. (a)

Kaufman, A. S. Verbal-performance IQ discrepancies on the WISC-R. Journal of Consulting and Clinical Psychology, 1976, 44, 739-744. (b)

Kaufman, A. S. The impact of WISC-R research for school psychologists. In C. R. Reynolds & T. B. Gutkin (Eds.), A handbook for school psychology. New York: John Wiley & Sons, 1981. (a)

Kaufman, A. S. The WISC-R and learning disabilities assessment: State of the art. Journal of Learning Disabilities, 1981, 14, 520-525. (b)

Kavale, K., & Nye, C. Identification criteria for learning disabilities: A survey of the research literature. Learning Disability Quarterly, 1981, 4, 383-388.

Kirk, S. A. Behavioral diagnosis and remediation of learning disabilities. Proceedings of the annual meeting of the conference on exploration into problems of the perceptually handicapped child, 1963, 1.

Kirk, S. A., & Bateman, B. Diagnosis and remediation of learning disabilities. Exceptional Children, 1962, 29, 73-78.

Kirk, S. A., & Elkins, J. Characteristics of children enrolled in the Child Service Demonstration Centers. Journal of Learning Disabilities, 1975, 8, 630-637.

Kretke, G. L.; Martin, L. P.; Averill, M.; & Hopkins, K. D. A district standardized testing program. Office of Evaluation and Guidance, Boulder Valley Public Schools, September 1976.

LaGrow, S. J., & Prochnow-LaGrow, J. E. Technical adequacy of the most popular tests selected by responding school psychologists in Illinois. Psychology in the Schools, 1982, 19, 186-189.

Larsen, S. C., & Hammill, D. D. The relationship of selected visual-perceptual abilities in school learning. Journal of Special Education, 1975, 9, 281-291.

Larsen, S. C.; Rogers, D.; & Sowell, V. The use of selected perceptual tests in differentiating between normal and learning disabled children. Journal of Learning Disabilities, 1976, 9, 85-90

Lumsden, J. Review of Illinois Test of Psycholinguistic Abilities, Revised Edition. In O.K. Buros (Ed.), The eighth mental measurement yearbook, vol I. Highland Park, N.J.: The Gryphon Press, 1978.

MacMillan, D. L., & Meyers, C. E. Educational labeling of handicapped learners. In D. C. Berliner (Ed.), Review of research in education, Vol. VII. Washington, D.C.: American Educational Research Association, 1979.

McDaniels, G. Assessing handicapped students: Beyond identification. In W. Schrader (Ed.), Measurement and educational policy: Proceedings of the 1978 ETS invitational conference. San Francisco: Jossey-Bass, 1979.

McLeod, J. Educational underachievement: Toward a defensible psychometric definition. Journal of Learning Disabilities, 1979, 12, 332-370.

Mehrens, W. A., & Lehmann, I. J. Measurement and evaluation in education and psychology (2nd ed.). New York: Holt, Rinehart and Winston, 1975.

Mercer, C. D.; Forgnone, C.; & Wolking, W. D. Definitions of learning disabilities used in the United States. Journal of Learning Disabilities, 1976, 9, 376-386.

Mercer, J. R. Protection in evaluation procedures. In PEP: Developing criteria for the evaluation of protection in evaluation procedures provisions. Philadelphia: Research for Better Schools, 1979.

National Advisory Committee on Handicapped Children. Special education for handicapped children. First Annual Report. Washington, D.C.: U.S. Department of Health, Education, and Welfare, January 31, 1968.

Newcomer, P. L., & Hammill, D. D. Psycholinguistics in the Schools. Columbus, Ohio: Charles E. Merrill Publishing, 1976.

Norman, C. A., & Zigmond, N. Characteristics of children labeled and served as learning disabled in school systems affiliated with Child Service Demonstration Centers. Journal of Learning Disabilities, 1980, 13, 16-21.

Olson, J. L., & Mealor, D. J. Learning disabilities identification: Do researchers have the answer? Learning Disability Quarterly, 1981, 4, 389-392.

Poland, S.; Ysseldyke, J.; Thurlow, M.; & Mirkin, P. Current assessment and decision-making practices in school settings as reported by directors of special education. Research Report No. 14. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Poplin, M. S. The severely learning disabled: Neglected or forgotten? Learning Disability Quarterly, 1981, 4, 330-335.

Reynolds, C. R., & Gutkin, T. B. Test scatter on the WPPSI: Normative analyses of the standardization sample. Journal of Learning Disabilities, 1981, 14, 460-464.

Sabatino, D. A., & Miller, T. L. The dilemma of diagnosis in learning disabilities: Problems and potential directions. Psychology in the Schools, 1980, 17, 76-86.

Salvia, J., & Ysseldyke, J. E. Assessment in special and remedial education. Boston: Houghton-Mifflin, 1978.

Schumaker, J. B.; Deshler, D. D.; Alley, G. R.; & Warner, M. M. An epidemiological study of learning disabled adolescents in secondary schools: The relationship of family factors to the condition of learning disabilities. Research Report No. 17. Lawrence, Kansas: The University of Kansas Institute for Research in Learning Disabilities, 1980.

Senf, G. M. Implications of the final procedures for evaluating specific learning disabilities. Journal of Learning Disabilities, 1978, 11, 124-126.

Shepard, L. An evaluation of the regression discrepancy method for identifying children with learning disabilities. Journal of Special Education, 1980, 14, 79-80.

Shepard, L. A., & Smith, M. L. An evaluation of the identification of children with learning disabilities in Colorado. Learning Disability Quarterly, in press.

Shepard, L., & Smith, M. L.; with Davis, A.; Glass, G. V; Riley, A.; & Vojir, C. Evaluation of the identification of perceptual-communicative disorders in Colorado. Laboratory of Educational Research, University of Colorado, Boulder, Colorado 80309, February 1981. ED 216 037.

Smith, M. L. How educators decide who is learning disabled. Springfield, Ill.: Charles C. Thomas, 1982.

Thorndike, R. L., & Hagen, E. P. Measurement and evaluation in psychology (4th ed.). New York: John Wiley & Sons, 1977.

Thurlow, M. T. Views of placement team members subsequent to their participation in meetings. In J. Ysseldyke, B. Algozzine, & M. Thurlow (Eds.), A Naturalistic investigation of special education team meetings. Research Report No. 40. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled. Research Report No. 11. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Tucker, J. A. Ethnic proportions in classes for the learning disabled: Issues in nonbiased assessment. Journal of Special Education, 1980, 14 93-105.

United States Office of Education. Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. Federal Register, 1977, 42, 65082-65085.

Warner, M. M.; Schumaker, J. B.; Alley, G. R.; & Deshler, D. D. Learning disabled adolescents in the public schools: Are they different from other low achievers? EEQ/Adolescents and Young Adults. 1980.

Weener, P. D., & Senf, G. M. Learning disabilities. In H.E. Mitzel, J. H. Best, & W. Rabinowitz (Eds.), Encyclopedia of educational research (5th ed.). New York: The Free Press, 1982. Pp. 1059-1068.

Wright, D. Evaluation of simulated decision making: A response to Ysseldyke, Algozzine, Regan and Potter. Psychology in the Schools, 1980, 17, 541-542.

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art. Research Report No 1. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1977.

Ysseldyke, J. E. Implementing the protection in evaluation procedures provision of P.L. 94-142. In PEP: Developing criteria for the evaluation of protection in evaluating procedures provisions. Philadelphia: Research for Better Schools, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information. Journal of Special Education, 1981, 15, 429-435.

Ysseldyke, J.; Algozzine, B.; Regan, R.; & McGue, M. The influence of test scores and naturally occurring pupil characteristics on psychoeducational decision making with children. Journal of School Psychology, 1981, 19, 167-177.

Ysseldyke, J.; Algozzine, B.; Regan, R.; & Potter, M. On "unenlightening data" and "bogus problems": A response to Dan Wright. Psychology in the Schools, 1980, 17, 543-544. (a)

Ysseldyke, J. E.; Algozzine, B.; Regan, R.; & Potter, M. Technical adequacy of tests used by professionals in simulated decision making. Psychology in the Schools, 1980, 17, 202-209. (b)

Ysseldyke, J. E.; Algozzine, B.; Shinn, M. R.; & McGue, M. Similarities and differences between low achievers and students classified learning disabled. The Journal of Special Education, 1982, 16, 73-85.

Appendix A

COMPUTATION OF STATISTICALLY SIGNIFICANT DISCREPANCIES*

EXAMPLE

A fourth grade boy was referred for special education assessment in November (4.2). He had been referred once before in second grade and his third grade teacher had reported that he "continues to be behind in reading and math." His WISC-R scores were 95 for both verbal and performance, resulting in a full scale IQ of 94. On the PIAT his achievement levels were: Math 3.0, Reading Recognition 3.9, and Reading Comprehension 3.5. Are these scores, especially the low math performance, significantly discrepant from ability reflected in the 94 IQ?

SIMPLE z SCORE OR PERCENTILE COMPARISONS

	Scores	Corresponding Percentiles	z Standard Scores
IQ	94	34 %ile (obtained by looking up z score in normal curve table)	$z = \frac{\text{Score} - \bar{X}}{s} = \frac{94-100}{15} = -.4$
PIAT Math	3.0	21 %ile (obtained by reference to tables in test manual)	z= .80 (obtained by looking up 21 %ile in normal curve table)

*Some familiarity is assumed with standard deviations (s), with correlation coefficients (r) and with standard scores (z) based on the normal distribution. These concepts are taught in introductory statistics or measurement courses. Readers who wish to review this material briefly should see Chapter 2 in Hopkins, K. D. & Stanley, J. C., Educational and Psychological Measurement and Evaluation, Sixth Edition. Prentice Hall: Englewood Cliffs, New Jersey, 1981.

As shown above, an IQ score of 94 is at the 34th percentile. Because percentile equivalents are not usually given in IQ test manuals, the percentile must be determined by first computing the standard z score and then looking up the z score in a normal curve table. Knowing that the WISC-R has a mean of 100 and a standard deviation of 15, the z score can be computed or a conversion table can be developed as in Table 1.

For achievement tests, percentile equivalents are usually given in norms tables in the test manuals. The percentile rank (for children in the first third of the fourth grade year, e.g. 4.0 - 4.3) must then be converted to a standard score by referencing a normal curve table. (Table 1 can serve as an abbreviated version; more complete tables are found in most measurement and statistics textbooks.) Standard z scores will be needed in the next section to determine the significance of the discrepancy.

For our example, the percentile rank of the achievement score, the 21st percentile, can be compared roughly with the ability percentile of 34. The math percentile is below the ability ranking, but is the difference significant? Unfortunately this question cannot be answered in terms of percentile ranks because a difference of 13 points (34-21) has different meaning at different points on the scale. Therefore, differences must be tested using standard scores. In the above example, the difference between ability and achievement is .4 z score units (-.4 -(-.8)).

STATISTICALLY SIGNIFICANT DISCREPANCIES

Small differences between intelligence and achievement are normal. Small discrepancies might be caused by normal developmental differences, subtle differences in opportunity to learn, and lack of

Table 1

Percentile Ranks and Standard Scores for Deviation IQs (σ = 15)

IQ Score	Percentile Rank	Standard z score	IQ Score	Percentile Rank	Standard z score
145	99.87	3.00	99	47	.07
144	99.83	2.93	98	45	.13
143	99.79	2.87	97	42	.20
142	99.74	2.80	96	39	.27
141	99.69	2.73	95	37	.33
140	99.62	2.67	94	34	.40
139	99.53	2.60	93	32	.47
138	99	2.53	92	30	.53
137	99	2.47	91	27	.60
136	99	2.40	90	25	.67
135	99	2.33	89	23	.73
134	99	2.27	88	21	.80
133	99	2.20	87	19	.87
132	98	2.13	86	18	.93
131	98	2.07	85	16	1.00
130	98	2.00	84	14	1.07
129	97	1.93	83	13	1.13
128	97	1.87	82	12	1.20
127	96	1.80	81	10	1.27
126	96	1.73	80	9	1.33
125	95	1.67	79	8	1.40
124	95	1.60	78	7	1.47
123	94	1.53	77	6	1.53
122	93	1.47	76	5	1.60
121	92	1.40	75	5	1.67
120	91	1.33	74	4	1.73
119	90	1.27	73	4	1.80
118	88	1.20	72	3	1.87
117	87	1.13	71	3	1.93
116	86	1.07	70	2	2.00
115	84	1.00	69	2	2.07
114	82	.93	68	2	2.13
113	81	.87	67	1	2.20
112	79	.80	66	1	2.27
111	77	.73	65	1	2.33
110	75	.67	64	1	2.40
109	73	.60	63	1	2.47
108	70	.53	62	1	2.53
107	68	.47	61	.47	2.60
106	66	.40	60	.38	2.67
105	63	.33	59	.31	2.73
104	61	.27	58	.26	2.80
103	58	.20	57	.21	2.87
102	55	.13	56	.17	2.93
101	53	.07	55	.13	3.00
100	50	.00			

Note: These conversions can be used for any test with a mean of 100 and a standard deviation of 15.

perfect correlation between abilities and specific learning tasks. Small or moderate discrepancies can also be the result of measurement error. Therefore, before an observed difference between ability and achievement test scores can be considered severe, it must be, at a minimum, a reliable difference. That is, the discrepancy must not be due to chance; it must not be due to random errors in one or both tests. The formulae explained below are used to compute statistically significant discrepancies between IQ and achievement test results. These formulae reflect how big a difference is likely to occur just by chance. Discrepancies, then, which are significant are those which would occur only rarely by chance and are therefore more likely to be real differences.

To actually compute a discrepancy between IQ and achievement, test scores must be converted to a common metric or scale. If tests have different means and standard deviations, comparisons of raw scores are meaningless. Usually, as in the example above, scores should be standardized to the z score scale ($\mu = 0$, $\sigma = 1$) using either the norm group mean and standard deviation or percentile conversions from the normal distribution. If the IQ and achievement tests happen to already be in the same metric, as with the WISC-R and WRAT ($\mu = 100$, $\sigma = 15$), conversion to the z score scale is not necessary, so long as one remembers to use the appropriate standard deviation ($\sigma = 15$) in the formula given below.

To test whether a discrepancy (such as a z difference of .4) is large enough to be reliable, the standard error of the difference (SEM_{dif}) has to be calculated for the particular pair of tests, e.g. the WISC-R IQ test and PIAT math test. The standard error of the difference is dependent on both the reliability of the difference (r_{dif}) and the standard deviation of the difference (s_{dif}). Therefore,

these values were calculated first. The necessary formulae and a computational example using WISC-R Full Scale IQ and PIAT Math are given below. The derivation of these formulae is further explained in Salvia and Ysseldyke (1978) and Thorndike and Hagen (1977).

PREREQUISITE DATA OBTAINED FROM TEST MANUALS*

Subscripts are used to denote theWISC-R Full Scale IQ as test 1 and PIAT Math as test 2.

Standard deviation, $s_1 = 1$ (in z units)
$s_2 = 1$ (in z units)

Reliability, $r_{11} = .95$ WISC-R manual pp. 32-33 test-retest correlations averaged across ages.

$r_{22} = .74$ Math PIAT manual: test-retest correlations, median across grades of within grade correlations.

Between-test correlation, $r_{12} = .53$ Math PIAT manual: median across grades correlation of PIAT subtest with PPVT. This is conservative estimate since WISC-R is more reliable than PPVT.

*Wechsler, D. Wechsler Intelligence Scale for Children-Revised: Manual. New York: Psychological Corporation, 1974.

Dunn, L. M. & Markwardt, F. C. Peabody Individual Achievement Test: Manual. Circle Pines, Minn.: American Guidance Services, 1970.

Reliability of the difference

$$r_{dif} = \frac{\frac{1}{2}(r_{11} + r_{22}) - r_{12}}{1 - r_{12}}$$

$$= \frac{\frac{1}{2}(.95 + .74) - .53}{.47} = \frac{.315}{.47} = .67 \text{ WISC-R \& PIAT Math}$$

Standard deviation of the difference

$$s_{dif} = \sqrt{s_1^2 + s_2^2 - 2r_{12}s_1s_2}$$

$$= \sqrt{1 + 1 - 2(.53)} = \sqrt{2-1.06} = \sqrt{.94} = .97 \text{ WISC-R \& PIAT Math}$$

Standard error of the difference

$$SEM_{dif} = s_{dif}\sqrt{1 - r_{dif}}$$

$$= .97\sqrt{1-.67} = .97(.574) = .56 \text{ WISC-R \& PIAT Math}$$

In the Shepard and Smith (1981) studies, standard errors of the differences were computed for several of the most frequently used pairs of tests. These standard errors are reported in Table 2. The appropriate standard error of the difference for comparing WISC-R IQ scores with PIAT math scores was computed above to be .56. This value is found in Table 2 by reading in the WISC-R row under the column for PIAT math. Similarly, if the Woodcock Reading Test had been used with the WISC-R, the standard error of the difference from Table 2 would be .39. Often school district test specialists can be asked to develop tables similar to Table 2 for frequently used tests. Thus, it is possible to avoid the fairly elaborate computations given above once the appropriate tables are available.

COMPLETION OF EXAMPLES

So far we know that the difference between IQ and math achievement for this fourth grade boy is .4 in z score units and that the appropriate standard error for judging this difference is .56. The only remaining requirement is to reference the correct probability distribution used to establish statistical significance.

Table 2

Standard Errors of the Difference for Most Frequently Used Pairs of Tests (in z standard score units)

Most Frequently Used IQ Tests	Most Frequently Used Achievement Tests							
	WRAT		PIAT		Woodcock Reading	CTBS typical of group norm-referenced tests		
	Reading	Math	Read. Rec.	Math		Reading	Math	
Peabody Picture Vocabulary Test (PPVT)	.57	.616	.584	.70	.57	.54	.54	
	(8.54)	(9.22)						
	(on IQ scale)							
WISC-R	.514	.558	.40	.56	.39	.33	.33	
	(7.718)	(8.358)						
	(on IQ scale)							

The sampling distribution for difference scores (discrepancies) is a normal distribution. Therefore the probability statements derived from the normal probability density function can be used to determine how large a difference is significantly greater than chance.

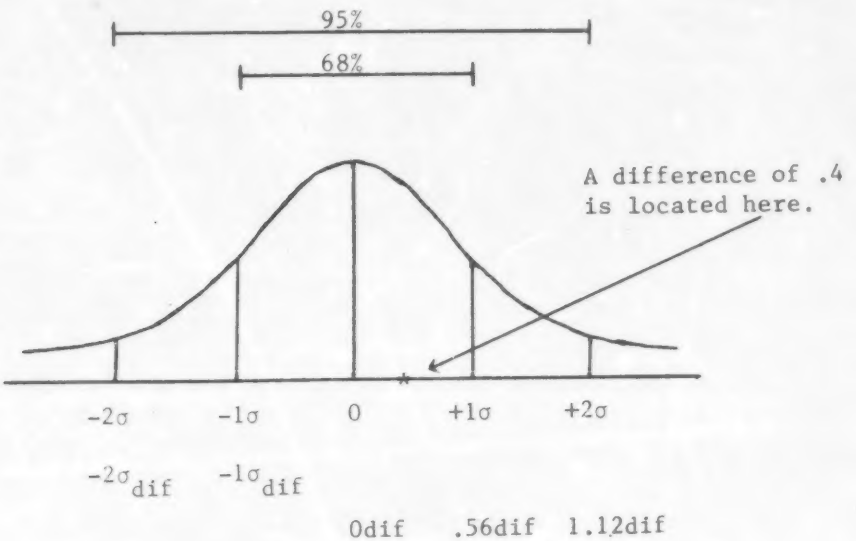


Figure. Areas (probabilities) under the normal curve for $\pm 1\sigma$ and $\pm 2\sigma$

In our example, the observed difference, .4, is less than one standard error of the difference (.4 < .56). Therefore, we know from the statistical model that differences this large would occur in more than two-thirds of the cases just by chance. The lower math score in this case is neither unusual nor "significantly" different from the IQ score. In fact, the difference would have to be more than twice as large, roughly 1.12 in standard score units, before it could be considered significant at the .05 level.

ERIC/TM Report 84

ASSESSMENT OF LEARNING DISABILITIES

by

Lorrie A. Shepard

University of Colorado

The purpose of this report is to present a summary of the issues in learning disability (LD) assessment. How do educators determine who is learning disabled? What practices are recommended? The main focus of the paper is on specific, relatively technical points that influence the validity of assessment. A basic premise is, however, that technical concerns are only one of the factors influencing the validity of placements.

Therefore, the paper is organized into two major sections: (1) the context of LD identification and (2) technical issues in LD assessment. In the first section, specific propositions regarding the context of LD identification are advanced with supporting evidence. In the second section, recommendations are made for the improved training or retraining of specialist. The recommendations include contextual changes that are likely to help clinicians be willing to make more rigorous diagnoses.

ORDER FORM

Please send_____copies of ERIC/TM Report 84, "Assessment of Learning Disabilities," at $6.50 per copy.

Name_____

Address_____

_____Zip_____

TOTAL ENCLOSED $_____

Return this form to:
ERIC/TM
Educational Testing Service
Princeton, NJ 08541

RECENT TITLES

IN THE ERIC/TM REPORT SERIES

#83 - Statistical Methodology in Meta-Analysis, by Larry V. Hedges. 12/82 $7.00.

#82 - Microcomputers in Educational Research, by Craig W. Johnson. 12/82 $8.50.

#81 - A Bibliography to Accompany the Joint Committee's Standards on Educational Evaluation, compiled by Barbara M. Wildemuth. 107p. $8.50.

#80 - The Evaluation of College Remedial Programs, by Jeffrey K. Smith and others. 12/81. $8.50.

#79 - An Introduction to Rasch's Measurement Model, by Jan-Eric Gustafsson. 12/81, $5.50.

#78 - How Attitudes Are Measured: A Review of Investigations of Professional, Peer, and Parent Attitudes toward the Handicapped, by Marcia D. Horne. 12/80, $5.50.

#77 - The Reviewing Processes in Social Science Publications: A Review of Research, by Susan E. Hensley, and Carnot E. Nelson. 12/80, $4.00.

#76 - Intelligence Testing, Education, and Chicanos: An Essay in Social Inequality, by Adalberto Acquirre Jr. 12/80, $5.50.

#75 - Contract Grading, by Hugh Taylor. 12/80, $7.50.

#74 - Intelligence, Intelligence Testing and School Practices, by Richard DeLisi. 12/80, $4.50.

#73 - Measuring Attitudes Toward Reading, by Ira Epstein. 12/80, $9.50.

#72 - Methods of Identifying Gifted Minority Students, by Ernest M. Bernal. 12/80, $4.50.

#71 - Sex Bias in Testing: An Annotated Bibliography, by Barbara Hunt. 12/79, $5.00.

#70 - The Role of Measurement in the Process of Instruction, by Jeffrey K. Smith. 12/79, $3.50.

#68 - The Educational Implications of Piaget's Theory and Assessment Techniques, by Richard DeLisi. 11/79, $5.00.

#66 - Competency-Based Graduation Requirements: A Point of View, by Mary Ann Bunda. 1978, $2.00.

#65 - The Practice of Evaluation, by Clare Rose and Glenn F. Nyre. 12/77, $5.00.

#63 - Perspectives on Mastery Learning & Mastery Testing, by Jeffrey K. Smith. 1977, $3.00.